



# ***STIC Search Report***

## ***Biotech-Chem Library***

**STIC Database Tracking Number: 99145**

**TO: Sheridan Swope**  
**Location: cm1/12D12/10D01**  
**Art Unit: 1652**  
**Friday, July 18, 2003**

**Case Serial Number: 09/966880**

**From: Barb O'Bryen**  
**Location: Biotech-Chem Library**  
**CM1-6A05**  
**Phone: 308-4291** *1327*

**barbara.obryen@uspto.gov**

### **Search Notes**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 17, 2003, 21:03:07 ; Search time 3550 seconds  
(without alignments)  
12856.023 Million cell updates/sec

Title: US-09-966-880A-7  
Perfect score: 2818  
Sequence: 1 agagaccatcattatga.....aaaaaaaaaaaaaaaaaaaaa 2818

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 500 summaries

Database :

EST:\*

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estin:\*
- 4: em\_estmu:\*
- 5: em\_estov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_htc:\*
- 9: gb\_est1:\*
- 10: gb\_est2:\*
- 11: gb\_htc:\*
- 12: gb\_est3:\*
- 13: gb\_est4:\*
- 14: gb\_est5:\*
- 15: em\_estfun:\*
- 16: em\_estom:\*
- 17: gb\_gss:\*
- 18: em\_gss\_hum:\*
- 19: em\_gss\_inv:\*
- 20: em\_gss\_pln:\*
- 21: em\_gss\_vrt:\*
- 22: em\_gss\_fun:\*
- 23: em\_gss\_nam:\*
- 24: em\_gss\_mus:\*
- 25: em\_gss\_other:\*
- 26: em\_gss\_pro:\*
- 27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	777	27.6	953	14	BQ065440
2	760	27.0	856	9	AL559877
3	750	26.6	872	12	BG758510
4	730	25.9	743	12	BG686133
5	718	25.5	1052	14	BQ055935
6	663	23.5	664	12	BG754140

7	608	21.6	820	12	BG755526
8	606	21.5	820	12	BG757089
9	592	21.0	693	12	BF975096
10	558	19.8	570	12	BG341546
11	540	19.2	541	12	BF238155
12	523	18.6	570	10	AW504807
13	497	17.6	693	12	BG757392
14	456	16.2	889	12	BG686876
15	448	15.9	604	10	AW978582
16	448	15.9	604	10	AW978582
17	442	15.7	442	12	BF975166
18	380	13.5	511	12	AG1016902
19	369	13.1	374	10	BG170756
20	353	12.5	353	9	AA831307
21	349	12.4	476	10	AW135547
22	346	12.3	464	10	AW452648
23	321	11.4	321	9	AI468242
24	312	11.1	413	10	AW768364
25	308	10.9	363	9	AI249134
26	297	10.5	956	12	BF664355
27	296	10.5	422	10	AW263139
28	296	10.5	528	9	AA954956
29	281	10.0	332	9	AA504653
30	279	9.9	928	12	BG398364
31	278	9.9	454	10	AW134750
32	207	7.3	939	12	BG341819
33	205	7.3	517	12	BG170824
34	199	7.1	215	10	AW444889
35	180	6.4	463	9	AA521498
36	153	5.4	507	9	AL581406
37	147	5.2	268	9	AA879422
38	101	3.6	1140	12	BF664352
39	85	3.0	364	10	AW182255
40	73	2.6	154	9	AA074342
41	73	2.6	878	12	BE988834
42	72	2.6	586	10	AV719004
43	70	2.5	252	10	BE008586
44	70	2.5	253	10	BE012103
45	70	2.5	318	9	AA548692
46	70	2.5	334	10	AV759459
47	70	2.5	346	9	AA211885
48	70	2.5	346	10	AV647070
49	70	2.5	437	17	AQ056722
50	70	2.5	448	14	N22395
51	70	2.5	544	17	AQ284662
52	70	2.5	642	10	AW973278
53	70	2.5	649	17	AQ741937
54	70	2.5	705	17	AG185819
55	69	2.4	571	17	BI4085
56	67	2.4	300	17	AQ276581
57	67	2.4	397	17	AZ303190
58	67	2.4	483	13	BI023361
59	67	2.4	725	17	AQ478179
60	67	2.4	841	17	AQ750497
61	66	2.3	286	10	AW882256
62	66	2.3	286	10	AW882261
63	66	2.3	286	10	AW882265
64	66	2.3	286	10	AW882267
65	66	2.3	388	10	AW518030
66	66	2.3	395	17	B56190
67	66	2.3	402	9	AA679794
68	66	2.3	414	17	B72122
69	66	2.3	430	17	AQ111323
70	66	2.3	434	17	AQ878262
71	66	2.3	451	9	AI653493
72	66	2.3	526	10	BE301068
73	65	2.3	395	10	BE156416
74	65	2.3	400	10	BE156526
75	65	2.3	443	12	BE940056
76	65	2.3	573	17	AQ318054
77	64	2.3	468	12	BG059756
78	64	2.3	470	17	AQ344499
79	64	2.3	479	14	N49425

BG755526	602716206
BG757089	602715124
BF975096	60245679
BG341546	602463652
BF238155	601811880
AW504807	UI-HF-BN0
BG757392	602711022
BG686876	602650861
AW978582	EST390691
BF975166	602444657
AG1016902	ou313003.x
BG170756	602323871
AW401901	UI-HF-BK0
AA831307	OC70A06.s
AW135547	UI-H-B11-
AW452648	UI-H-B13-
AI468242	tg35905.x
AW768364	hk73d04.x
AI249134	qh79c07.x
BF664355	602146013
AW263139	xn78d03.x
AA954956	qp24a12.s
AA504653	aa63f11.s
BG398364	602439832
AW134750	UI-H-B11-
BG341819	602463552
BG170824	602323763
AW444889	UI-H-B13-
AA521498	aa73h06.s
AL581406	AL581406
AA879422	OJ91C11.s
BF664352	602146010
AW182255	x172f10.x
AA074342	em15903.s
BE988834	601682124
AV719004	AV719004
BE008586	RC5-BN015
BE012103	RC5-BN105
AA548692	nk03g01.s
AV759459	AV759459
AA211885	z190c03.s
AV647070	AV647070
AQ056722	CIT-HSP-2
N22395	yw37f08.s1
AQ284662	RPC111-78
AW973278	EST385376
AQ741937	HS_5569_B
AG185819	Pan trogl
BI4085	A-837A4.TP
AQ276581	CITBI-EI-
AZ303190	UP_436-15
BI023361	CM4-MT024
AQ478179	RPC1-11-2
AQ750497	HS_5576_B
AW882256	RC5-OT005
AW882261	RC5-OT005
AW882265	RC5-OT005
AW882267	RC5-OT005
AW518030	xx65h01.x
B56190	CIT-HSP-200
AA679794	ag92b12.s
B72122	CIT978SK-A-
AQ111323	CIT-HSP-2
AQ878262	HS_2272_B
AI653493	tg95a01.x
BE301068	ba84g06.x
BE156416	QVO-HT036
BE156526	QVO-HT036
BE940056	RC3-UT003
AQ318054	RPC111-11
BG059756	na153c05.
AQ344499	RPC111-13
N49425	yv21c11.r1

```

80      64      2.3      657      17      A0480012      A0480012      RPCI-11-2      c 153      57      2.0      475      10      A0008184
c 81      64      2.3      665      17      AG041398      Pan trogl      154      57      2.0      484      17      B43436
c 82      64      2.2      263      17      A2516708      RPCI-11-1      c 155      57      2.0      488      13      B1052695
c 83      63      2.2      380      12      BF933309      ILS-NT027      c 156      57      2.0      495      9      A1471740
c 84      63      2.2      424      9      A1673749      tw9e11.x      c 157      57      2.0      510      10      A0962508
c 85      63      2.2      445      17      A0238522      RPCI11-70      c 158      57      2.0      512      12      BG460639
c 86      63      2.2      454      17      A0268452      RPCI11-69      c 159      57      2.0      521      9      A1890840
c 87      63      2.2      490      17      A0017563      CIT-HSP-2      c 160      57      2.0      524      17      B90841
c 88      63      2.2      509      17      A0624960      CITBI-E1-      c 161      57      2.0      529      17      A0532786
c 89      63      2.2      639      17      AG127767      Pan trogl      c 162      57      2.0      531      14      BM843015
c 90      63      2.2      683      9      AL704146      DKFp686D      c 163      57      2.0      533      12      BG431768
c 91      63      2.2      716      9      AL042927      DKFp434K      c 164      57      2.0      538      17      A0471842
c 92      63      2.2      734      14      BQ707295      AGENCOURT      165      57      2.0      553      13      B1014701
c 93      62      1034      12      BG755005      602711511      c 166      57      2.0      565      11      BC020924
c 94      61      2.2      308      9      AA993041      ot92f06.s      c 167      57      2.0      569      14      BQ417358
c 95      61      2.2      396      14      BM988047      UI-H-DF0-      c 168      57      2.0      578      12      BF675962
c 96      61      2.2      400      9      AA838140      oel1b04.s      c 169      57      2.0      589      12      BF130687
c 97      61      2.2      404      17      A0596495      HS-5191_A      c 170      57      2.0      592      14      BQ287900
c 98      61      2.2      430      17      A0414005      RPCI-11-1      c 171      57      2.0      622      17      AG123931
c 99      61      2.2      460      17      A0537631      RPCI-11-3      c 172      57      2.0      654      17      AG067831
c 100      61      2.2      498      17      A0551954      RPCI-11-4      c 173      57      2.0      659      10      BM876459
c 101      61      2.2      504      9      AA577824      nn24e06.s      c 174      57      2.0      665      14      BQ954941
c 102      61      2.2      512      14      BQ029024      UI-H-DF0-      c 175      57      2.0      666      17      AG042343
c 103      61      2.2      539      9      AU144268      AU144268      c 176      57      2.0      668      17      AG138481
c 104      61      2.2      562      17      A0792364      HS-5255_B      c 177      57      2.0      669      17      AG147516
c 105      60      2.1      228      17      AQ115117      RPCI11-50      c 178      57      2.0      670      17      AG086214
c 106      60      2.1      347      9      AA565232      nj23c04.s      c 179      57      2.0      678      17      AZ852685
c 107      60      2.1      397      10      AW572202      xk92a08.x      c 180      57      2.0      701      10      AW876458
c 108      60      2.1      442      9      AA137013      zn97b08.r      c 181      57      2.0      707      17      AG019375
c 109      60      2.1      650      17      AG045269      Pan trogl      c 182      57      2.0      725      12      BG433250
c 110      60      2.1      674      17      AG067832      Pan trogl      c 183      57      2.0      756      14      BQ429325
c 111      60      2.1      843      12      BG758815      602713177      c 184      57      2.0      766      14      BQ429325
c 112      59      2.1      335      9      AA594694      no01g01.s      c 185      57      2.0      772      17      AQ898994
c 113      59      2.1      383      17      B50561      B50561      c 186      57      2.0      882      9      AL133861
c 114      59      2.1      390      9      AA533762      nj3b03.s      c 187      57      2.0      927      14      BQ718159
c 115      59      2.1      419      9      A1818505      wk73e12.x      c 188      57      2.0      1298      11      AF113689
c 116      59      2.1      511      17      AQ173021      HS-3200_B      c 189      57      2.0      2971      17      AF101960
c 117      59      2.1      569      9      AL048969      DKFp434M      c 190      57      2.0      259      9      AA988569
c 118      59      2.1      619      17      AQ001362      CIT-HSP-2      c 191      57      2.0      56      2.0      299      10      AW270451
c 119      58      2.1      194      9      A1832674      at71b02.x      c 192      56      2.0      333      10      BE155011
c 120      58      2.1      440      13      B1037013      CM1-NT020      c 193      56      2.0      338      17      B85513
c 121      58      2.1      557      17      AQ265532      RPCI11-68      c 194      56      2.0      437      17      AQ774179
c 122      58      2.1      677      17      AG081343      Pan trogl      c 195      56      2.0      507      9      AG601712
c 123      58      2.1      697      17      AG092595      Pan trogl      c 196      56      2.0      657      17      AG149297
c 124      58      2.1      701      17      AG151917      Pan trogl      c 197      56      2.0      658      17      AG079497
c 125      57      2.0      211      9      AA708889      zf94c03.s      c 198      56      2.0      668      17      HSA408806
c 126      57      2.0      249      17      B76008      B76008      c 199      56      2.0      668      17      HSA408806
c 127      57      2.0      266      9      AA847504      oel1b03.s      c 200      56      2.0      658      17      HSA408806
c 128      57      2.0      290      9      AA376856      EST9327      c 201      55      2.0      658      17      HSA408806
c 129      57      2.0      296      9      AA522486      n169e08.s      c 202      55      2.0      332      12      BE777611
c 130      57      2.0      313      13      B1062421      IL3-UT011      c 203      55      2.0      357      9      AA581895
c 131      57      2.0      326      9      AA342238      EST47717      c 204      55      2.0      359      17      AQ007622
c 132      57      2.0      347      13      B1493542      B1493542      c 205      55      2.0      381      10      AW272389
c 133      57      2.0      369      9      AA329995      EST33665      c 206      55      2.0      381      10      AW272389
c 134      57      2.0      382      17      B79812      CIT-HSP-205      c 207      55      2.0      411      9      A1805350
c 135      57      2.0      385      17      B51502      CIT978SK-A-      c 208      55      2.0      418      9      A1805355
c 136      57      2.0      396      14      BQ023082      UI-1-BB1p      c 209      55      2.0      430      17      B38671
c 137      57      2.0      400      9      A1014283      am50h08.s      c 210      55      2.0      457      17      AQ531017
c 138      57      2.0      410      17      B82099      RPCI11-25I3      c 211      55      2.0      474      12      BF218498
c 139      57      2.0      413      9      AL047349      DKFp586G      c 212      55      2.0      476      14      W60853
c 140      57      2.0      421      17      B79819      CIT-HSP-205      c 213      55      2.0      497      10      AW814044
c 141      57      2.0      430      10      AW233836      xp24f02.x      c 214      55      2.0      501      17      AQ588985
c 142      57      2.0      432      14      H73970      yu04g03.s1      c 215      55      2.0      505      12      BG496448
c 143      57      2.0      432      17      BH634644      TAR-CH5-N      c 216      55      2.0      569      17      AQ484672
c 144      57      2.0      435      17      AQ029116      RPCI11-39      c 217      55      2.0      611      17      AQ486033
c 145      57      2.0      452      12      BF725761      bx19e11.y      c 218      55      2.0      617      9      A0206340
c 146      57      2.0      457      17      A0759119      HS-2249_B      c 219      55      2.0      621      14      BM703459
c 147      57      2.0      453      17      AQ894865      HS-2036_B      c 220      55      2.0      628      13      BM310792
c 148      57      2.0      454      10      A0718485      AV718485      c 221      55      2.0      630      10      BE066054
c 149      57      2.0      459      17      AQ561531      HS-5197_A      c 222      55      2.0      640      12      BF325535
c 150      57      2.0      471      14      BQ035433      HS-2050_B      c 223      55      2.0      664      10      BE065974
c 151      57      2.0      472      17      BQ030345      UI-H-DT0-      c 224      55      2.0      685      17      AG141466
c 152      57      2.0      473      17      B43705      HS-1058-A1-      c 225      55      2.0      692      10      BE065971

```



226	55	2.0	708	17	AG11119	AG11119 Pan trogl	299	53	1.9	583	10	BE078293	BE078293 CM0-BT061
227	55	2.0	750	17	AQ745035	AQ745035 HS_5506_A	300	53	1.9	617	17	AQ488260	AQ488260 RPCI-11-2
C 228	55	2.0	1069	13	BM464113	BM464113 AGENCOURT	C 301	53	1.9	628	12	BG433506	BG433506 602497939
C 229	54	1.9	179	9	AA578326	AA578326 n138g04.s	302	53	1.9	632	12	BE902017	BE902017 601676824
C 230	54	1.9	219	9	AA484658	AA484658 ne93g08.s	303	53	1.9	668	17	AQ505772	AQ505772 RPCI-11-2
C 231	54	1.9	251	17	AQ474929	AQ474929 CITBI-El-	304	53	1.9	674	17	AG132336	AG132336 Pan trogl
C 232	54	1.9	259	17	B75724	B75724 RPT111-12C1	305	53	1.9	691	10	AV682637	AV682637 AV682637
C 233	54	1.9	273	9	A1758981	A1758981 ty96b01.x	306	53	1.9	693	17	AG116744	AG116744 Pan trogl
C 234	54	1.9	300	17	AQ473028	AQ473028 CITBI-El-	307	53	1.9	1093	13	BM469364	BM469364 AGENCOURT
C 235	54	1.9	312	9	AA573067	AA573067 nm34g08.s	308	52	1.8	174	9	AA369200	AA369200 EST80751
C 236	54	1.9	346	9	AA662598	AA662598 nr17c11.s	309	52	1.8	200	14	H67663	H67663 yr76b07.s1
C 237	54	1.9	356	9	A1277454	A1277454 qm84d07.x	310	52	1.8	221	14	H67708	H67708 yr72b07.s1
C 238	54	1.9	377	17	AQ166130	AQ166130 HS_3071_A	311	52	1.8	245	12	BF987027	BF987027 RC1-GN002
C 239	54	1.9	389	9	A1244127	A1244127 qv90a07.x	312	52	1.8	250	9	AA318652	AA318652 EST20754
C 240	54	1.9	400	9	A1679082	A1679082 tu61a02.x	313	52	1.8	262	9	AA365347	AA365347 EST76110
C 241	54	1.9	408	17	AZ518006	AZ518006 RPCI-11-2	314	52	1.8	305	9	AA303040	AA303040 EST114070
C 242	54	1.9	420	17	AQ036829	AQ036829 CIT-HSP-2	315	52	1.8	313	9	AA297443	AA297443 EST113044
C 243	54	1.9	446	9	AA427860	AA427860 zw50a11.s	C 316	52	1.8	313	12	BF17893	BF17893 7p85g10.x
C 244	54	1.9	450	17	AQ338985	AQ338985 HS_5021_A	C 317	52	1.8	345	9	A1689008	A1689008 tx10e07.x
C 245	54	1.9	490	17	AQ034701	AQ034701 CIT-HSP-2	C 318	52	1.8	362	9	A1049634	A1049634 an31e04.x
C 246	54	1.9	548	12	BG898990	BG898990 HOA4-1-F-	C 319	52	1.8	389	12	BG060038	BG060038 naf47b03.
C 247	54	1.9	565	9	AL599722	AL599722 DFE2p13H	C 320	52	1.8	413	17	AQ033882	AQ033882 HS_2189_B
C 248	54	1.9	584	17	AQ030963	AQ030963 CITBI-El-	C 321	52	1.8	414	9	A1807372	A1807372 wf47d07.x
C 249	54	1.9	589	12	BF853273	BF853273 MR2-EN009	C 322	52	1.8	423	10	AW963497	AW963497 EST375570
C 250	54	1.9	649	17	AG054765	AG054765 Pan trogl	C 323	52	1.8	426	10	AW467233	AW467233 h607h01.x
C 251	54	1.9	669	17	AG157455	AG157455 Pan trogl	C 324	52	1.8	445	14	R88750	R88750 yp93e01.s1
C 252	54	1.9	692	17	AG104929	AG104929 Pan trogl	C 325	52	1.8	457	9	AA543015	AA543015 ni55e05.s
C 253	54	1.9	707	17	AG180372	AG180372 Pan trogl	C 326	52	1.8	465	14	N95099	N95099 za28a02.r1
C 254	54	1.9	720	17	AV762203	AV762203 AV762203	C 327	52	1.8	466	17	AQ412292	AQ412292 RPCI-11-1
C 255	54	1.9	801	17	AQ744761	AQ744761 HS_5506_A	C 328	52	1.8	497	17	AQ419955	AQ419955 RPCI-11-2
C 256	54	1.9	816	17	AQ988085	AQ988085 HS_3122_B	C 329	52	1.8	511	17	BH609775	BH609775 HIV12H08
C 257	54	1.9	861	14	BQ231417	BQ231417 AGENCOURT	C 330	52	1.8	521	17	AQ431260	AQ431260 HS_5120_A
C 258	54	1.9	966	14	BQ876394	BQ876394 AGENCOURT	C 331	52	1.8	525	17	AQ394682	AQ394682 CITBI-El-
C 259	54	1.9	5468	17	AQ839834	AQ839834 260L13-C5	C 332	52	1.8	530	17	AQ357292	AQ357292 CITBI-El-
C 260	53	1.9	183	9	AA021326	AA021326 ze69c11.r	C 333	52	1.8	535	17	AQ269088	AQ269088 RPCI11-74
C 261	53	1.9	203	13	BI015299	BI015299 PM0-ET025	C 334	52	1.8	551	17	AQ266700	AQ266700 RPCI11-74
C 262	53	1.9	220	13	BI058821	BI058821 RC4-GN007	C 335	52	1.8	552	10	BE300871	BE300871 ba78c05.x
C 263	53	1.9	234	9	AA158288	AA158288 zo76f06.s	C 336	52	1.8	578	17	AQ267476	AQ267476 RPCI11-75
C 264	53	1.9	270	9	AA323439	AA323439 EST26216	C 337	52	1.8	590	9	AU127566	AU127566 AU127566
C 265	53	1.9	276	10	AW833285	AW833285 RC0-TT000	C 338	52	1.8	626	12	BE890239	BE890239 601508330
C 266	53	1.9	298	14	F00214	F00214 HSB80D051.S	C 339	52	1.8	635	12	BF901984	BF901984 PM4-MT020
C 267	53	1.9	299	9	AA377102	AA377102 EST89626	C 340	52	1.8	656	12	BF820327	BF820327 CM0-RT020
C 268	53	1.9	306	10	AW275788	AW275788 xp55b12.x	C 341	52	1.8	690	17	AG112514	AG112514 Pan trogl
C 269	53	1.9	333	9	AA468247	AA468247 nc76g04.s	C 342	52	1.8	695	17	AG085567	AG085567 Pan trogl
C 270	53	1.9	339	12	BF751955	BF751955 RC4-BN053	C 343	52	1.8	710	14	BQ014503	BQ014503 UI-H-ED1-
C 271	53	1.9	352	12	BF751953	BF751953 RC4-BN053	C 344	52	1.8	729	10	BE270175	BE270175 601185983
C 272	53	1.9	365	12	BF751949	BF751949 RC4-BN053	C 345	52	1.8	762	13	BI910261	BI910261 603069482
C 273	53	1.9	375	9	AA828970	AA828970 od81f04.s	C 346	52	1.8	821	12	BE967682	BE967682 601648988
C 274	53	1.9	375	10	AW833297	AW833297 RC0-TT000	C 347	52	1.8	877	14	BQ230629	BQ230629 AGENCOURT
C 275	53	1.9	379	12	BF739035	BF739035 PM4-KT004	C 348	52	1.8	889	14	BQ214673	BQ214673 AGENCOURT
C 276	53	1.9	387	10	BE157400	BE157400 RC0-HT037	C 349	52	1.8	935	12	BG432843	BG432843 602496051
C 277	53	1.9	387	10	BE157414	BE157414 RC0-HT037	C 350	52	1.8	1003	12	BG026261	BG026261 602291610
C 278	53	1.9	389	9	AA018517	AA018517 ze50f05.r	C 351	52	1.8	1238	13	BM455282	BM455282 AGENCOURT
C 279	53	1.9	389	10	BE350475	BE350475 h14c01.x	C 352	52	1.8	1875	13	BI005195	BI005195 Homo sap1
C 280	53	1.9	390	14	H86665	H86665 yt02d02.r1	C 353	52	1.8	175	9	AA153656	AA153656 tq46d10.x
C 281	53	1.9	395	17	AQ127179	AQ127179 HS_3042_B	C 354	51	1.8	177	9	AA094109	AA094109 c11865.se
C 282	53	1.9	404	13	BI037268	BI037268 MR4-TT014	C 355	51	1.8	186	12	BF215231	BF215231 601846224
C 283	53	1.9	408	9	AA468318	AA468318 nc77a04.s	C 356	51	1.8	195	12	BG058240	BG058240 nah22e10.S
C 284	53	1.9	413	9	AA055987	AA055987 zf55d06.r	C 357	51	1.8	214	14	F00818	F00818 HSB66E102.S
C 285	53	1.9	413	9	AT571161	AT571161 tn85h04.x	C 358	51	1.8	219	9	AA224338	AA224338 nr14e05.s
C 286	53	1.9	416	10	AW341978	AW341978 hb74d10.x	C 359	51	1.8	223	12	BG231402	BG231402 na142b09.
C 287	53	1.9	416	13	BI035522	BI035522 PM3-GN037	C 360	51	1.8	231	14	R92608	R92608 yq07g05.s1
C 288	53	1.9	465	9	A1336026	A1336026 qt42h03.x	C 361	51	1.8	236	10	AW854958	AW854958 PM0-CF026
C 289	53	1.9	466	17	AQ413749	AQ413749 RPCI-11-2	C 362	51	1.8	238	17	AQ505893	AQ505893 RPCI-11-2
C 290	53	1.9	473	9	AA478515	AA478515 zw95c12.s	C 363	51	1.8	240	14	H56912	H56912 yr07d08.s1
C 291	53	1.9	478	17	AQ545518	AQ545518 CITBI-El-	C 364	51	1.8	235	14	F13678	F13678 HSF1H115.SC
C 292	53	1.9	484	10	AW405593	AW405593 UI-HF-BK0	C 365	51	1.8	236	10	AW854958	AW854958 PM0-CF026
C 293	53	1.9	488	14	N52274	N52274 yv47e09.s1	C 366	51	1.8	240	14	H56912	H56912 yr07d08.s1
C 294	53	1.9	512	17	AQ586138	AQ586138 RPCI-11-4	C 367	51	1.8	242	9	A1307579	A1307579 tb35h05.x
C 295	53	1.9	513	17	AQ682298	AQ682298 HS_5500_B	C 368	51	1.8	245	9	A1682665	A1682665 wc64e10.x
C 296	53	1.9	516	13	BG995743	BG995743 MR2-HT16	C 369	51	1.8	246	9	A1378730	A1378730 tb91h11.x
C 297	53	1.9	521	12	BF906164	BF906164 IL3-MT026	C 370	51	1.8	247	9	AT382324	AT382324 ta71d11.x
C 298	53	1.9	528	17	AQ726110	AQ726110 HS_5403_B	C 371	51	1.8	247	9	AT382324	AT382324 ta71d11.x

372	51	1.8	250	17	AZ755597	ex01h05.x	C 445	51	1.8	378	9	AA525753	AA525753 n157a04.s
373	51	1.8	253	12	BF854778	RC6-FN008	446	51	1.8	378	14	BQ337978	BQ337978 QV2-NN107
374	51	1.8	259	12	BI015027	PM0-ET025	C 447	51	1.8	381	17	AO374918	AO374918 RPI11-15
375	51	1.8	265	12	BF811807	MR2-CT018	C 448	51	1.8	384	12	BF742112	BF742112 CMI-HB000
376	51	1.8	266	10	AA314494	EST18633	C 449	51	1.8	385	14	BQ693971	BQ693971 1000114.H
377	51	1.8	266	10	AW079664	x559g12.x	C 450	51	1.8	387	9	AO1061313	AO1061313 an32q07.x
378	51	1.8	269	12	BF854664	MR2-EN009	C 451	51	1.8	390	9	AA229935	AA229935 nc51q10.r
379	51	1.8	269	17	AO040838	CIT-HSP-2	C 452	51	1.8	390	10	AW847291	AW847291 RC0-CT020
380	51	1.8	271	9	AA947269	ok20a01.s	C 453	51	1.8	391	13	BG942512	BG942512 ax26b07.x
381	51	1.8	272	17	BH141089	UF-511-21	C 454	51	1.8	393	9	AI276985	AI276985 qm59a06.x
382	51	1.8	279	17	AZ755629	ev02c07.x	C 455	51	1.8	404	9	AA553497	AA553497 nk81a05.s
383	51	1.8	280	9	AI814739	wk65b06.x	C 456	51	1.8	407	9	AI753113	AI753113 zq08c11.x
384	51	1.8	282	14	F32705	HSPD25675.H	C 457	51	1.8	416	9	AA194633	AA194633 cr06f04.s
385	51	1.8	283	12	BF680898	602156158	458	51	1.8	416	17	B79034	B79034 CIT-HSP-629
386	51	1.8	284	14	BM710652	UI-E-DX0-	459	51	1.8	419	9	AI357052	AI357052 qx62c09.x
387	51	1.8	288	9	AA569474	nm38f11.s	460	51	1.8	420	12	BE745183	BE745183 601573920
388	51	1.8	288	10	BE003190	QV4-BN009	C 461	51	1.8	420	14	BQ582255	BQ582255 l114q08.x
389	51	1.8	292	9	AA603530	np15b10.s	462	51	1.8	421	10	AW963470	AW963470 EST375543
390	51	1.8	294	10	BE065600	RC3-BT031	463	51	1.8	423	9	AA776665	AA776665 ac79h07.s
391	51	1.8	301	14	H69351	yu19d12.s1	C 464	51	1.8	426	17	AO118480	AO118480 HS-3011.B
392	51	1.8	303	9	AI865807	wk86c05.x	465	51	1.8	430	12	BF873332	BF873332 IL5-ET011
393	51	1.8	303	9	AA601376	no16d02.s	C 466	51	1.8	430	12	BF928492	BF928492 IL2-NT020
394	51	1.8	303	10	AW881075	QV0-OT003	C 467	51	1.8	433	17	AO046754	AO046754 RPI111-35
395	51	1.8	303	17	AO490385	RPI1-11-2	C 468	51	1.8	434	14	BQ710197	BQ710197 AGNCNCOURT
396	51	1.8	307	12	BF849267	CM3-EN007	C 469	51	1.8	435	9	AI829331	AI829331 wk60f09.x
397	51	1.8	307	17	B57499	CIT-HSP-200	C 470	51	1.8	435	9	AA581471	AA581471 nd17f107.s
398	51	1.8	308	9	AA715075	nv50h02.r	471	51	1.8	435	12	BF773305	BF773305 CM3-IT004
399	51	1.8	308	12	BF976855	602146501	472	51	1.8	435	12	BG469693	BG469693 602534014
400	51	1.8	309	10	BE246822	TCBAP1E51	C 473	51	1.8	436	10	AW104748	AW104748 xd56e07.x
401	51	1.8	310	10	AW068316	cn20e11.y	474	51	1.8	437	9	AL712763	AL712763 DKF7p686p
402	51	1.8	313	12	BF676352	602084530	C 475	51	1.8	437	14	R48424	R48424 yj64f05.r1
403	51	1.8	314	17	B90405	CIT-HSP-216	C 476	51	1.8	438	14	W42997	W42997 zc24d11.s1
404	51	1.8	315	9	AA608751	ae56h07.s	C 477	51	1.8	440	17	AO089740	AO089740 HS-3006.B
405	51	1.8	319	14	F29968	HSPD20151.H	C 478	51	1.8	444	12	BF026590	BF026590 601672723
406	51	1.8	321	12	BG117013	602315864	C 479	51	1.8	445	9	AI769271	AI769271 wq36a02.x
407	51	1.8	330	9	AI285651	qu80d11.x	C 480	51	1.8	445	17	B43746	B43746 HS-1058-A2-
408	51	1.8	333	14	T98037	ye30c05.s1	C 481	51	1.8	446	10	BE302658	BE302658 ba73q03.y
409	51	1.8	334	9	AA715173	nv52q08.r	C 482	51	1.8	446	14	BH833654	BH833654 K-IST0108
410	51	1.8	335	10	AW815249	QV4-ST021	C 483	51	1.8	449	9	AA599011	AA599011 ac41e10.s
411	51	1.8	336	17	AZ755602	ev01h10.x	C 484	51	1.8	449	17	AO805344	AO805344 HS-3214.A
412	51	1.8	339	9	AA639946	np12e05.s	C 485	51	1.8	450	14	BM745900	BM745900 K-IST0019
413	51	1.8	339	17	AZ756263	ev03h02.r	C 486	51	1.8	450	17	B88194	B88194 CIT-HSP-219
414	51	1.8	339	17	AZ756706	ev15a05.r	C 487	51	1.8	455	10	BE409402	BE409402 601303664
415	51	1.8	340	10	AW855736	RC1-CT027	C 488	51	1.8	456	17	AO225744	AO225744 HS-2005.B
416	51	1.8	340	17	AZ756412	ev07d10.r	C 489	51	1.8	457	9	AL134278	AL134278 DKF7p547H
417	51	1.8	343	17	AZ756878	ew02g12.x	C 490	51	1.8	457	10	BE390587	BE390587 601284940
418	51	1.8	345	9	AA693484	ah20f10.s	C 491	51	1.8	458	9	AI288696	AI288696 qm06a08.x
419	51	1.8	346	9	AI446601	tj25a11.x	C 492	51	1.8	458	10	AW073871	AW073871 xb04c06.x
420	51	1.8	351	13	BM353189	ig76g10.x	C 493	51	1.8	458	14	R60189	R60189 yhl2h11.s1
421	51	1.8	352	9	AI611533	tt65f07.x	C 494	51	1.8	459	17	B41911	B41911 HS-1054-B1-
422	51	1.8	352	9	AI612070	tt64a06.x	C 495	51	1.8	460	9	AI361063	AI361063 qy04a03.x
423	51	1.8	352	12	BF900022	CM3-MT019	C 496	51	1.8	463	9	AI332680	AI332680 qq33q07.x
424	51	1.8	353	9	AA507837	ng86d05.s	C 497	51	1.8	468	14	N63746	N63746 za18e11.s1
425	51	1.8	355	12	BG152682	na758h12.	C 498	51	1.8	469	12	BF229305	BF229305 PM4-AN009
426	51	1.8	361	9	AA779075	zj42f01.s	499	51	1.8	469	12	BF229310	BF229310 PM4-AN009
427	51	1.8	362	9	AI266003	qm46d10.x	500	51	1.8	470	17	AO562723	AO562723 HS-2079.B
428	51	1.8	362	9	AA251477	zs08d02.r							
429	51	1.8	363	9	AI367721	qv75g11.x							
430	51	1.8	363	10	AW166611	xf64g09.x							
431	51	1.8	364	9	AI205077	an04f08.x							
432	51	1.8	364	9	AI205128	an04h08.x							
433	51	1.8	365	9	AA837817	oe39h03.s							
434	51	1.8	365	10	AW752493	RC0-CT020							
435	51	1.8	365	12	BF848354	PM4-EN006							
436	51	1.8	365	17	AO085904	HS-2164.B							
437	51	1.8	366	9	AI283938	qt71e11.x							
438	51	1.8	367	10	AV760014	AV760014							
439	51	1.8	367	17	AQ239048	RPI111-66							
440	51	1.8	370	14	M77888	EST01472.Fe							
441	51	1.8	372	12	BF822429	MRI-RT002							
442	51	1.8	373	14	F03490	HSC12A052.n							
443	51	1.8	374	9	AI023070	ow65e09.s							
444	51	1.8	376	9	AA579179	nf28d09.s							

RESULT 1.  
BQ065440  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
BQ065440  
AGENCOURT\_6855061 NIH\_MGC\_99 Homo sapiens cDNA clone IMAGE:5929977  
5', mRNA sequence.  
BQ065440  
BQ065440.1 GI:19894486  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

## ALIGNMENTS

```

REFERENCE 1 (bases 1 to 953)
AUTHORS NIH-MGC http://mgs.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Lou Staudt
cDNA Library preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM2108 row: p column: 10
High quality sequence stop: 634.
FEATURES             Location/Qualifiers
     source            1..953
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone_lib="NIH_MGC_99"
                        /tissue_type="lymphoma, cell line"
                        /lab_host="DH10B (phage-resistant)"
                        /note="Organ: lymph; Vector: pOTB7; Site_1: XhoI; Site_2:
                        EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                        into EcoRI/XhoI sites using the following 5' adaptor:
                        GGCACGAG(G). Size-selected >500bp for average insert size
                        1.8kb. Library constructed by Ling Hong in the laboratory
                        of Gerald M. Rubin (University of California, Berkeley)
                        using ZAP-cDNA synthesis kit (Stratagene) and Superscript
                        II RT (Life Technologies). Note: this is a NIH_MGC
                        Library."
BASE COUNT      238 a   236 c   233 g   246 t

Query Match      27.6%; Score 777; DB 14; Length 953;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 777; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 TTGAAGTGAAGATTTTCTGGCTGAGACTTGCAGGAGGCAAGAAGACACTCTGGACACC 76
DB 1 TTGAAGTGAAGATTTTCTGGCTGAGACTTGCAGGAGGCAAGAAGACACTCTGGACACC 60
QY 77 ACTATGACAGCCTCTTGATGAACCGGAGAAAGTTCTTTTACCAATTCAAAAATGCCGC 136
DB 61 ACTATGACAGCCTCTTGATGAACCGGAGAAAGTTCTTTTACCAATTCAAAAATGCCGC 120
QY 137 TGGGCTAAGGTCGGCGTCAGACCTACCTGTGCTAGCTAGTGAAGAGCGGTGACAGTGT 196
DB 121 TGGGCTAAGGTCGGCGTCAGACCTACCTGTGCTAGCTAGTGAAGAGCGGTGACAGTGT 180
QY 197 ACATCCTTTTCACTGGACTTTGGTTATCTTCGCAATPAAGACGGCTGCCACGTGGAATG 256
DB 181 ACATCCTTTTCACTGGACTTTGGTTATCTTCGCAATPAAGACGGCTGCCACGTGGAATG 240
QY 257 CTCCTCCTCGGTACATCTCGGACTGGGACCTAGACCCCTGGCGGCTGTACCGCGTCACC 316
DB 241 CTCCTCCTCGGTACATCTCGGACTGGGACCTAGACCCCTGGCGGCTGTACCGCGTCACC 300
QY 317 TGGTTCACTCTCGGAGCCCTGCTACGACTGTGCCGACATGTGCCGCACTTCTGCCA 376
DB 301 TGGTTCACTCTCGGAGCCCTGCTACGACTGTGCCGACATGTGCCGCACTTCTGCCA 360
QY 377 GGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCCG 436
DB 361 GGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCCG 420
QY 437 AGGCTGACCCGAGGGGCTCGGGGCTGCACCGCGCGGGTGCAATAGCCATCATG 496
DB 421 AAGGCTGAGCCGAGGGGCTCGGGGCTGCACCGCGCGGGTGCAATAGCCATCATG 480
QY 497 ACCTTCAAGAGATTTATTTTACTGCTGGAATACCTTTGTAGAAAACCATGAAGAACTTTC 556

```

```

DB 481 ACCTTCAAAAGATTTATTTTACTGCTGGAATACCTTTTGTAGAAAACCATGAAGAAGCTTTC 540
QY 557 AAAGCCTGGGAAGGGCTGCATGAAATTCAGTTTCGCTCTCCACAGACAGCTTCGCGGCATC 616
DB 541 AAAGCCTGGGAAGGGCTGCATGAAATTCAGTTTCGCTCTCCACAGACAGCTTCGCGGCATC 600
QY 617 CTTTTCGCCCTGTATGAGGTTGATGACTTACGAGACGCATTTCTGTAATTTGGGACTTTGA 676
DB 601 CTTTTCGCCCTGTATGAGGTTGATGACTTACGAGACGCATTTCTGTAATTTGGGACTTTGA 660
QY 677 TAGCAACTTCCAGGAATGCACACAGCATGAAATATCTCTGCTGAAGACAGTGCATATAA 736
DB 661 TAGCAACTTCCAGGAATGCACACAGCATGAAATATCTCTGCTGAAGACAGTGCATATAA 720
QY 737 AACAGTCCCTCAAGTCTTCTCTGTTTTTATTTCAACTCTCACTTCTTAGAGTTT 793
DB 721 AACAGTCCCTCAAGTCTTCTCTGTTTTTATTTCAACTCTCACTTCTTAGAGTTT 777

RESULT 2
AL559877      856 bp      mRNA      linear      EST 16-FEB-2001
LOCUS      AL559877 LTI_FL011_BC1 Homo. sapiens cDNA clone CS0DG003YB14 5 prime
DEFINITION      , mRNA sequence.
ACCESSION      AL559877
VERSION      AL559877.1 GI:12905793
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 856)
AUTHORS      Li, W.-B., Gruber, C., Jessee, J., and Polayes, D.
TITLE      Full-length cDNA libraries and normalization
JOURNAL      Unpublished (2001)
COMMENT      Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES             Location/Qualifiers
     source            1..856
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone="CS0DG003YB14"
                        /clone_lib="LTI_FL011_BC1"
                        /sex="male"
                        /tissue_type="B cells from Burkitt lymphoma"
                        /lab_host="DH10B"
                        /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
                        with a NotI-oligo(dT) primer. Five prime end enriched,
                        double-stranded cDNA was digested with Not I and cloned
                        into the Not I and Eco RV sites of the pCMVSPORT 6 vector.
                        Library was constructed by Life Technologies. Contact :
                        Peng Liang Life Technologies, a division of Invitrogen
                        9800 Medical Center Drive Rockville, Maryland 20850, USA
                        Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL :
                        http://fulllength.invitrogen.com"
BASE COUNT      209 a   217 c   202 g   226 t   2 others
ORIGIN

Query Match      27.0%; Score 760; DB 9; Length 856;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 810; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 62 GACACTCTGGACACCACTATGACAGCCCTTTGTATGAACCGGAGGAAGTTCTTTTACCAC 121
DB 1 GACACTCTGGACACCACTATGACAGCCCTTTGTATGAACCGGAGGAAGTTCTTTTACCAC 60
QY 122 TTCAAAAATGTCCGTGGGCTAAGGTCGGCTGAGACCTACTGTGCTACGTAGTGAAG 181
DB 61 TTCAAAAATGTCCGTGGGCTAAGGTCGGCTGAGACCTACTGTGCTACGTAGTGAAG 120
QY 182 AGGCGTGACAGTGCTACATCTTTTCACTGAGACTTTTGGTTATCTTCGCAATAAGAACGCG 241

```

```

121 AGGCGTGACAGTCTACATCTTTTCACTGGACTTTGGTTATCTTGCATAAAGACGGC 180
242 TGCCACGTGGAATTGCTCTTCCTCGCTACATCTCGGACTGGGACCTAGACCTGCCGC 301
181 TGCCACGTGGAATTGCTCTTCCTCCGCTACATCTCGGACTGGGACCTAGACCTGCCGC 240
302 TGCTACCGGCTCACCTGGTTTCACTCTCGAGGCCCTGCTAGGACTGTGCCGACATGTG 361
241 TGCTACCGGCTCACCTGGTTTCACTCTCGAGGCCCTGCTAGGACTGTGCCGACATGTG 300
362 GCGGACTTTCTCGGAGGGAACCCCAACCTCAGTCTGAGGATCTTCAACCGCGCCTCTAC 421
301 GCGGACTTTCTCGGAGGGAACCCCAACCTCAGTCTGAGGATCTTCAACCGCGCCTCTAC 360
422 TTCTGTGAGACCGCGAAGGCTGAGCCGAGGGCTGCGGCGCTGCACCGCGCGGGG 481
361 TTCTGTGAGACCGCGAAGGCTGAGCCGAGGGCTGCGGCGCTGCACCGCGCGGGG 420
482 CAATAGCCATCATGACCTTCAAAGATTATTTTACTGTGGAATCTTTGTAGAAAAC 541
421 CAATAGCCATCATGACCTTCAAAGATTATTTTACTGTGGAATCTTTGTAGAAAAC 480
542 CATGAAAGAACTTCAAAGCCTGGGAAGGCTGCATGAAAAATTCAGTTGCTCTCCAGA 601
481 CATGAAAGAACTTCAAAGCCTGGGAAGGCTGCATGAAAAATTCAGTTGCTCTCCAGA 540
602 CAGCTTCGCGGATCCTTTTGGCCCTGTATGAGTTGATGACTTACGAGACGATTCGT 561
541 CAGCTTCGCGGATCCTTTTGGCCCTGTATGAGTTGATGACTTACGAGACGATTCGT 600
662 ACTTTGGGACTTTGATAGCAACTTCCAGGAATGTCCACACGATGAAATATCTCTGCTGA 721
601 ACTTTGGGACTTTGATAGCAACTTCCAGGAATGTCCACACGATGAAATATCTCTGCTGA 660
722 ACACAGTGATAAAAACAGTCTTCAAGTCTCTCTGTTTATTTTCACTCTCACTCTCACT 781
661 AGACAGTGGATAAAAAAGTCCCTTCAAGTCTCTCTGTTTATTTTCACTCTCACTCTCACT 720
782 TTCTTAGAGTTTACAGAAAAATATTTATATACGACTCTTTAAAAAGATCTATGCTTGA 841
721 TTCTTAGAGTTTACAGAAAAATATTTATATACGACTCTTTAAAAAGATCTATGCTTGA 780
842 AAATAGAGAAGAACACAGGCTGCGCCAGG 872
781 AAATAGAGAAGAACACAGGCTGCGCCAGG 811
```

```

RESULT 3
LOCUS BG758510
DEFINITION 602712721F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4853069 5',
mRNA sequence.
ACCESSION BG758510
VERSION BG758510.1 GI:14069163
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 872)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
```

```

Plate: LLM1698 row: i column: 06
High quality sequence stop: 836.
FEATURES
source
1..872
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4853069"
/clone_lib="NIH_MGC_48"
/tissue_type="primary B-cells from tonsils (cell line)"
/lab_host="DH10B (phage-resistant)"
/note="Organ: B-cells; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT 211 a 221 c 212 g 228 t
ORIGIN
Query Match 26.6%; Score 750; DB 12; Length 872;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 800; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 53 AGGCAAGAGACACTCTGGACACCACCTATGGACAGCCTCTTTCATGAACCGGAGGAAGTTT 112
Db 39 AGGCAAGAGACACTCTGGACACCACCTATGGACAGCCTCTTTCATGAACCGGAGGAAGTTT 98
QY 113 CTTTACCAAATTCAAAATATGTCGCTGGGCTAAGGGTCGGCGTGAGACCTTACCTGTCTACTAC 172
Db 99 CTTTACCAAATTCAAAATATGTCGCTGGGCTAAGGGTCGGCGTGAGACCTTACCTGTCTACTAC 158
QY 173 GTAGTGAAGAGCGGTGACAGTCTACATCCTTTTTCACCTGGACCTTTGGTTATCTTCGCAAT 232
Db 159 GTAGTGAAGAGCGGTGACAGTCTACATCCTTTTTCACCTGGACCTTTGGTTATCTTCGCAAT 218
QY 233 AAGAACGGCTGCACGCTGGAATTTGCTTTCCTCCGCTACATCTCGGACTTGGACCTTAGAC 292
Db 219 AAGAAGGCTGCACGCTGGAATTTGCTTTCCTCCGCTACATCTCGGACTTGGACCTTAGAC 278
QY 293 CCGCGCGCTGCTACCGGCTCACCTGGTTTACCTCTCGAGGCCCTTGCCTACGACTGTGCG 352
Db 279 CCGCGCGCTGCTACCGGCTCACCTGGTTTACCTCTCGAGGCCCTTGCCTACGACTGTGCG 338
QY 353 CGACATGTGCGCGACTTTCTGCGAGGGAAACCCCAACCTCAGTCTGAGGATCTTCAACCGG 412
Db 339 CGACATGTGCGCGACTTTCTGCGAGGGAAACCCCAACCTCAGTCTGAGGATCTTCAACCGG 398
QY 413 CGCCTCTACTTCTGTGAGGACCGCAAGGCTGAGCCGAGGGGCTGCGGCGGCTGCACCGC 472
Db 399 CGCCTCTACTTCTGTGAGGACCGCAAGGCTGAGCCGAGGGGCTGCGGCGGCTGCACCGC 458
QY 473 GCGGGGTGCAAAATAGCCATCATGACCTTCAAAGATTATTTTACTGCTGGAATATCTTT 532
Db 459 GCGGGGTGCAAAATAGCCATCATGACCTTCAAAGATTATTTTACTGCTGGAATATCTTT 518
QY 533 GTAGAAAACCATGAAGAAGACTTTCAAGCCTGGGAGGCTGCATGAAATTCAGTTCT 592
Db 519 GTAGAAAACCATGAAGAAGACTTTCAAGCCTGGGAGGCTGCATGAAATTCAGTTCT 578
QY 593 CTCTCCACACAGCTTCGGCGCATCCTTTTGCCCTGTATGAGGTTGATGACTTACGAGAC 652
Db 579 CTCTCCACACAGCTTCGGCGCATCCTTTTGCCCTGTATGAGGTTGATGACTTACGAGAC 638
QY 653 GCATTTCTGACTTTTGGGACTTTTGTAGCAACTTCCAGGAATGTACACACATGATGAATAT 712
Db 639 GCATTTCTGACTTTTGGGACTTTTGTAGCAACTTCCAGGAATGTACACACATGATGAATAT 698
QY 713 CTCTCTCTGAAGACAGTGGATAAAAACAGTCTTCAAGTCTTCTCTGTTTATTTCTTCA 772
Db 699 CTCTCTCTGAAGACAGTGGATAAAAACAGTCTTCAAGTCTTCTCTGTTTATTTCTTCA 758
```

```

QY 773 ACTCTCACTTCTTAGAGTTTACAGAAAAATATTATATACGACTCTTTAAAAAGATCT 832
    |||||||
Db 759 ACTCTCACTTCTTAGAGTTTACAGCAAAATATTATATACGACTCTTTAAAAAGATCT 818
    |||||||
QY 833 ATGCTCTGAAATAGAGAAGG 853
    |||||||
Db 819 ATGCTCTGAAATAGAGAAGG 839

RESULT 4
BG686133
LOCUS 602638412F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4766234 5',
DEFINITION mRNA sequence.
ACCESSION BG686133
VERSION BG686133.1 GI:13917530
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 743)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1626 row: g column: 03
High quality sequence stop: 740.
Location/Qualifiers
1. 743
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4766234"
/clone_lib="NIH_MGC_48"
/tissue_type="primary B-cells from tonsils (cell line)"
/lab_host="DH10B (phage-resistant)"
/note="Organ: B-cells; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT 176 a 197 c 188 g 182 t
ORIGIN

Query Match 25.9%; Score 730; DB 12; Length 743;
Best Local Similarity 100.08; Pred. No. 0;
Matches 730; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GAACCATCATTAATTGAAGTGAGATTTTCTGGCCTGAGACTTGCGAGGGAGCGCAAGAAGA 63
    |||||||
Db 2 GAACCATCATTAATTGAAGTGAGATTTTCTGGCCTGAGACTTGCGAGGGAGCGCAAGAAGA 61
    |||||||
QY 64 CACTCTGCAGACCACCTATGGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTTACCAATT 123
    |||||||
Db 62 CACTCTGCAGACCACCTATGGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTTACCAATT 121
    |||||||
QY 124 CAAAATGTCCTGGCTAAGGGTCGGCGGTGAGACCTACCTGTCTACGTAGTGAAGAG 183
    |||||||
Db 122 CAAAATGTCCTGGCTAAGGGTCGGCGGTGAGACCTACCTGTCTACGTAGTGAAGAG 181
    |||||||

```

```

QY 184 GCGTGACAGTGCTACATCTCTTTTTCACGTGGACTTTGGTTATCTTCGCAATAAGAAGCGGTG 243
    |||||||
Db 182 GCGTGACAGTGCTACATCTCTTTTTCACGTGGACTTTGGTTATCTTCGCAATAAGAAGCGGTG 241
    |||||||
QY 244 CCACGTGGAATTGCTCTCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGCGCGGTG 303
    |||||||
Db 242 CCACGTGGAATTGCTCTCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGCGCGGTG 301
    |||||||
QY 304 CTACCGGCTCACCTGGTTTCACCTCTCTGGAGCCCTGCTTACGACTGTGCGCCGACATGTGGC 363
    |||||||
Db 302 CTACCGGCTCACCTGGTTTCACCTCTCTGGAGCCCTGCTTACGACTGTGCGCCGACATGTGGC 361
    |||||||
QY 364 CGACTTTTCTCGAGGGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGCCCTCTACTT 423
    |||||||
Db 362 CGACTTTTCTCGAGGGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGCCCTCTACTT 421
    |||||||
QY 424 CTGTGAGACCGCAAGGCTGAGCCCGAGGGGCTGGCGGGCTGCACCGCGCGGGGTGCA 483
    |||||||
Db 422 CTGTGAGACCGCAAGGCTGAGCCCGAGGGGCTGGCGGGCTGCACCGCGCGGGGTGCA 481
    |||||||
QY 484 AATAGCCATCATGACCTTCAAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACCA 543
    |||||||
Db 482 AATAGCCATCATGACCTTCAAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACCA 541
    |||||||
QY 544 TGAAGAAGAACTTTCAAAGCCTGGGAAGGCTGCATGAAATTCAGTTCGTCTCTCCAGACA 603
    |||||||
Db 542 TGAAGAAGAACTTTCAAAGCCTGGGAAGGCTGCATGAAATTCAGTTCGTCTCTCCAGACA 601
    |||||||
QY 604 GCTTCGGCGCATCTTTTGGCCCTGTATGAGGTTGATGACTTACGAGACGATTTTCGTAC 663
    |||||||
Db 602 GCTTCGGCGCATCTTTTGGCCCTGTATGAGGTTGATGACTTACGAGACGATTTTCGTAC 661
    |||||||
QY 664 TTTGGGACTTTGTAGCAACTTCCAGGAATGTCACACAGATGAAATATCTCTGCTGAAG 723
    |||||||
Db 662 TTTGGGACTTTGTAGCAACTTCCAGGAATGTCACACAGATGAAATATCTCTGCTGAAG 721
    |||||||
QY 724 ACAGTGGATA 733
    |||||||
Db 722 ACAGTGGATA 731

RESULT 5
BG055935
LOCUS AGENCOURT_6796291 NIH_MGC_99 Homo sapiens cDNA clone IMAGE:5808181
DEFINITION 5', mRNA sequence.
ACCESSION BG055935
VERSION BG055935.1 GI:19815262
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1052)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Lou Staudt
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2051 row: m column: 14
High quality sequence stop: 665.
Location/Qualifiers
1. 1052
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5808181"

```

FEATURES  
source

```

/clone_lib="NIH_MGC_99"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: lymph; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."
BASE COUNT      276 a      250 c      258 g      266 t      2 others
ORIGIN
Query Match      25.5%; Score 718; DB 14; Length 1052;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 718; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 17 TTGAAGTGAGATTTTCTGGCTGAGACTTGCAGGAGGCAAGAGACACTCTGGACACC 76
Db 1 TTGAAGTGAGATTTTCTGGCTGAGACTTGCAGGAGGCAAGAGACACTCTGGACACC 60
Qy 77 ACTATGGACAGCCTCTTGATGAACCGGAGGAGTTTCTTACCAATTCAAAAATCTCCG 136
Db 61 ACTATGGACAGCCTCTTGATGAACCGGAGGAGTTTCTTACCAATTCAAAAATCTCCG 120
Qy 137 TGGGCTAAGGTCGGCGTGAGACCTACCTGTGCTAGTGTAGTGAAGCGGTGACAGTGT 196
Db 121 TGGGCTAAGGTCGGCGTGAGACCTACCTGTGCTAGTGTAGTGAAGCGGTGACAGTGT 180
Qy 197 ACATCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGACGGCTGCCACGTGGAATTG 256
Db 181 ACATCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGACGGCTGCCACGTGGAATTG 240
Qy 257 CTCCTTCCTCGGTACATCTCGGACTTGGACCTAGACCTGGCGGCTGCTACCGGCTCACC 316
Db 241 CTCCTTCCTCGGTACATCTCGGACTTGGACCTAGACCTGGCGGCTGCTACCGGCTCACC 300
Qy 317 TGGTTCACTCTCTGGAGCCCTGCTACGACTGTGCGGACATGTGGCGGACTTCTCGGA 376
Db 301 TGGTTCACTCTCTGGAGCCCTGCTACGACTGTGCGGACATGTGGCGGACTTCTCGGA 360
Qy 377 GGGAAACCCCAACCTCAGTCTGAGGATCTTACCGGCGGCGCTACTCTGTGAGGACCGC 436
Db 361 GGGAAACCCCAACCTCAGTCTGAGGATCTTACCGGCGGCGCTACTCTGTGAGGACCGC 420
Qy 437 AAGGCTGAGCCGAGGGCTGCGGCGGCTGCACCGCGCGGGGTGCAATACCCATCATG 496
Db 421 AAGGCTGAGCCGAGGGCTGCGGCGGCTGCACCGCGCGGGGTGCAATACCCATCATG 480
Qy 497 ACCTTCAAGATTTATTTTACTGCTGGAAATCTTTGTAGAAAACCATGAAAGAACTTTC 556
Db 481 ACCTTCAAGATTTATTTTACTGCTGGAAATCTTTGTAGAAAACCATGAAAGAACTTTC 540
Qy 557 AAAGCCTGGGAAGGGCTGCATGAAATTCAGTTTCGCTCTCCAGACAGCTTCGGGCGATC 616
Db 541 AAAGCCTGGGAAGGGCTGCATGAAATTCAGTTTCGCTCTCCAGACAGCTTCGGGCGATC 600
Qy 617 CTTTTCGCCCTGTATGAGTTGATGACATTACGAGAGCAGATTTTCGTTGGGACTTTGA 676
Db 601 CTTTTCGCCCTGTATGAGTTGATGACATTACGAGAGCAGATTTTCGTTGGGACTTTGA 660
Qy 677 TAGCAACTTCCAGGAATGTCACACAGGATGAATATCTCTGCTGAAGACAGTGGATAA 734
Db 661 TAGCAACTTCCAGGAATGTCACACAGGATGAATATCTCTGCTGAAGACAGTGGATAA 718

```

```

RESULT 6
BG754140
LOCUS      602709681F1 NIH_MGC_48 664 bp mRNA linear EST 15-MAY-2001
DEFINITION mRNA sequence.

```

```

ACCESSION      BG754140
VERSION        BG754140.1
KEYWORDS       GI:14064793
SOURCE         human.
ORGANISM       Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 664)
AUTHORS        NIH-MGC http://mgc.nci.nih.gov/.
TITLE          National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL        Unpublished (1999)
COMMENT        Contact: Robert Strausberg, Ph.D.
                Email: cgapbs-remail.nih.gov
                Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
                cDNA Library Preparation: Ling Hong/Rubin Laboratory
                cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
                DNA Sequencing by: Incyte Genomics, Inc.
                Clone Distribution: MGC clone distribution information can be
                found through the I.M.A.G.E. Consortium/LLNL at:
                http://image.llnl.gov
                Plate: LCM1686 row: b column: 22
                High quality sequence stop: 662.
                Location/Qualifiers
                1..664
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:4846005"
                /clone_lib="NIH_MGC_48"
                /tissue_type="Primary B-cells from tonsils (cell line)"
                /lab_host="DH10B (phage-resistant)"
                /note="Organ: B-cells; Vector: pOTB7; Site_1: XhoI;
                Site_2: EcoRI; cDNA made by oligo-dT priming.
                Directionally cloned into EcoRI/XhoI sites using the
                following 5' adaptor: GGCAGGAG(G). Size-selected >500bp
                for average insert size 1.8kb. Library constructed by Ling
                Hong in the laboratory of Gerald M. Rubin (University of
                California, Berkeley) using ZAP-cDNA synthesis kit
                (Stratagene) and Superscript II RT (Life Technologies).
                Note: this is a NIH_MGC Library."
BASE COUNT      199 a      121 c      130 g      214 t
ORIGIN
Query Match      23.5%; Score 663; DB 12; Length 664;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 663; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 634 GGTTCATGACTTACCAGACGCAATTCGTACTTTGGGACTTTTGATAGCAACTTCCAGGAAT 693
Db 2 GGTTCATGACTTACCAGACGCAATTCGTACTTTGGGACTTTTGATAGCAACTTCCAGGAAT 61
Qy 694 GTCACACAGCATGAATATCTCTGCTGAAGACAGTGGATAAAAAACAGTCTTCAAGTCT 753
Db 62 GTCACACAGCATGAATATCTCTGCTGAAGACAGTGGATAAAAAACAGTCTTCAAGTCT 121
Qy 754 TCTCTGTTTTTATCTTCAACTCTCCTTCTTAGAGTTTACAGAAAAAATATTTATATA 813
Db 122 TCTCTGTTTTTATCTTCAACTCTCCTTCTTAGAGTTTACAGAAAAAATATTTATATA 181
Qy 814 GCACCTTTTAAAGACATCTATCTTTGAATAGAGGAACACAGGTCGCGCAGGCA 873
Db 182 GCACCTTTTAAAGACATCTATCTTTGAATAGAGGAACACAGGTCGCGCAGGCA 241
Qy 874 CGTGTGCAATTTGGTGCAGTTTGTGAATGCAACATTTGCCCTTACTGGGAATTAACAACT 933
Db 242 CGTGTGCAATTTGGTGCAGTTTGTGAATGCAACATTTGCCCTTACTGGGAATTAACAACT 301
Qy 934 GCAGGACCTGGGAGGATCCTAAAGTGTCAACCGTTTTTATGACATTTTATAGTAGATGAG 993
Db 302 GCAGGACCTGGGAGGATCCTAAAGTGTCAACCGTTTTTATGACATTTTATAGTAGATGAG 361
Qy 994 AGCAAGAGGTAGATCTCTTAAAAAGCATGGTGACAGGATCAAAATGTTTTTATATCAACATCC 1053
Db 362 AGCAAGAGGTAGATCTCTTAAAAAGCATGGTGACAGGATCAAAATGTTTTTATATCAACATCC 421

```

QY 1054 TTTATTATTGATTCATTTCAGTTAAACAGTGGTGTAGTGATAGATTTTTCATTCTTTT 1113  
|||||  
Db 422 TTTATTATTGATTCATTTCAGTTAAACAGTGGTGTAGTGATAGATTTTTCATTCTTTT 481  
QY 1114 CCCTGACGTTTACTTTCAAGTAACACAACTCTCCATCAGGCCATGATCATAGGACC 1173  
|||||  
Db 482 CCCTGACGTTTACTTTCAAGTAACACAACTCTCCATCAGGCCATGATCATAGGACC 541  
|||||  
QY 1174 TCCTAATGAGAGTATCTGGGTGATTGTGACCCCAACCATCTCTCCAAAGCAATTAATATC 1233  
|||||  
Db 542 TCCTAATGAGAGTATCTGGGTGATTGTGACCCCAACCATCTCTCCAAAGCAATTAATATC 601  
QY 1234 CAATCATGCGCTGTATGTTTTTAATCAGCAGAAGCATGTTTTTATGTTTGTACAAAAGAAG 1293  
|||||  
Db 602 CAATCATGCGCTGTATGTTTTTAATCAGCAGAAGCATGTTTTTATGTTTGTACAAAAGAAG 661  
QY 1294 ATT 1296  
|||  
Db 662 ATT 664  
  
RESULT 7  
BG755526 820 bp mRNA linear EST 15-MAY-2001  
LOCUS 602716206F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4856568 5',  
DEFINITION mRNA sequence.  
ACCESSION BG755526  
VERSION BG755526.1 GI:14066179  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 820)  
AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLCM1704 row: k column: 01  
High quality sequence stop: 638.  
Location/Qualifiers

FEATURES  
source  
1..820  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4856568"  
/clone\_lib="NIH\_MGC\_48"  
/tissue\_type="primary B-cells from tonsils (cell line)"  
/lab\_host="DH10B (phage-resistant)"  
/notes="Organ: B-cells; Vector: pOTB7; Site.1: XhoI;  
Site.2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCACGAG(G). Size-selected >500bp  
for average insert size 1.8kb. Library constructed by Ling  
Hong in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC library."  
BASE COUNT 245 a 161 c 168 g 246 t  
ORIGIN  
Query Match 21.6%; Score 608; DB 12; Length 820;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 608; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 665 TTGGGACTTTGTATGACCAACTCCAGGAATGTCCACACAGATGAAATATCTCTGCTGAAGA 724  
|||||  
Db 32 TTGGGACTTTGTATGACCAACTCCAGGAATGTCCACACAGATGAAATATCTCTGCTGAAGA 91  
|||||  
QY 725 CAGTGGATAAAAAACAGTCCCTTCAAGTCTTCTGTTTATTTATCTTCAACTCTCACTTTTC 784  
|||||  
Db 92 CAGTGGATAAAAAACAGTCCCTTCAAGTCTTCTGTTTATTTATCTTCAACTCTCACTTTTC 151  
|||||  
QY 785 TTAGAGTTTACAGAAAAAATATTTATATACGACTCTTTTAAAAAGATCTATGTCTTCAAAA 844  
|||||  
Db 152 TTAGAGTTTACAGAAAAAATATTTATATACGACTCTTTTAAAAAGATCTATGTCTTGAANA 211  
QY 845 TAGAGAAGGAACACAGGTCTGGCCAGGGACGTGTCGCAATTTGGTGAGTTTTGTAATGCAA 904  
|||||  
Db 212 TAGAGAAGGAACACAGGTCTGGCCAGGGACGTGTCGCAATTTGGTGAGTTTTGTAATGCAA 271  
QY 905 CATTGTCCCTACTCTGGGAATAACAGAACTGCAGGACCTGGGAGCATCCCTAAAAGTGTCAC 964  
|||||  
Db 272 CATTGTCCCTACTCTGGGAATAACAGAACTGCAGGACCTGGGAGCATCCCTAAAAGTGTCAC 331  
QY 965 GTTTTCTATGACTTTTAGGTAGGATGAGAGCAGAGGTAGATCTCTTAAAAAGCATGGTGA 1024  
|||||  
Db 332 GTTTTCTATGACTTTTAGGTAGGATGAGAGCAGAGGTAGATCTCTTAAAAAGCATGGTGA 391  
QY 1025 GAGGATCAAAATGTTTTTATATCAACATCCTTTTATTTATTTGATTCATTTGAGTTAAACAGTG 1084  
|||||  
Db 392 GAGGATCAAAATGTTTTTATATCAACATCCTTTTATTTGATTCATTTGAGTTAAACAGTG 451  
QY 1085 GTGTTAGTGATAGATTTTCTATTTCCCTTGACGTTTACTTTTCAAGTAAACAAAC 1144  
|||||  
Db 452 GTGTTAGTGATAGATTTTCTATTTCCCTTGACGTTTACTTTTCAAGTAAACAAAC 511  
QY 1145 TCTTCCATCAGGCCATGATCTATAGGACCTCTCTAATGAGAGATCTCTGGGTGATTGTGACC 1204  
|||||  
Db 512 TCTTCCATCAGGCCATGATCTATAGGACCTCTCTAATGAGAGATCTCTGGGTGATTGTGACC 571  
QY 1205 CCAAAACCATCTCTCCAAAGCATTAATATCAATCATGCGCTGTATGTTTTTAAATCAGCAGA 1264  
|||||  
Db 572 CCAAAACCATCTCTCCAAAGCATTAATATCAATCATGCGCTGTATGTTTTTAAATCAGCAGA 631  
QY 1265 AGCATGTTT 1272  
|||||  
Db 632 AGCATGTTT 639  
  
RESULT 8  
BG757089 820 bp mRNA linear EST 15-MAY-2001  
LOCUS 602715124F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4855517 5',  
DEFINITION mRNA sequence.  
ACCESSION BG757089  
VERSION BG757089.1 GI:14067742  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 820)  
AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLCM1704 row: o column: 06  
High quality sequence stop: 675.  
Location/Qualifiers

FEATURES

```
source
1. .820
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:485517"
/clone_lib="NIH_MGC_48"
/tissue_type="primary B-cells from tonsils (cell line)"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: B-cells; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
BASE COUNT 210 a 230 c 200 g 180 t
ORIGIN

Query Match 21.5%; Score 606; DB 12; Length 820;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 656; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGAAGACCATTAATGAAGTGAGATTTTCTGGCCTGAGACTTTCGAGGAGGCAAGA 60
Db
6 AGAAGACCATTAATGAAGTGAGATTTTCTGGCCTGAGACTTTCGAGGAGGCAAGA 65
QY 61 AGACACTCTGGACACCACTATGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCA 120
Db 66 AGACACTCTGGACACCACTATGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCA 125
QY 121 ATTCAAAATGTCGCTGGCTGAAGGTGCGGTGAGACTACTCTGTCTACGTAGTAA 180
Db 126 ATTCAAAATGTCGCTGGCTGAAGGTGCGGTGAGACTACTCTGTCTACGTAGTAA 185
QY 181 GAGCGGTGACAGTGCTACATCTTTTCACTGGACTTTTGGTTATCTTCGCAATAAGAACGG 240
Db 186 GAGCGGTGACAGTGCTACATCTTTTCACTGGACTTTTGGTTATCTTCGCAATAAGAACGG 245
QY 241 CTCGCCAGTGGAAATGCTTCTCCGCTACATCTCGGACTGGGACTAGACCTGGCCG 300
Db 246 CTCGCCAGTGGAAATGCTTCTCCGCTACATCTCGGACTGGGACTAGACCTGGCCG 305
QY 301 CTGCTACCGGCTCACCTGGTTACCTCCTGGAGCCCTGTACGACTGTGCCCGACATGT 360
Db 306 CTGCTACCGGCTCACCTGGTTACCTCCTGGAGCCCTGTACGACTGTGCCCGACATGT 365
QY 361 GGCCGACTTTCTCGGAGGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCTCTTA 420
Db 366 GGCCGACTTTCTCGGAGGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCTCTTA 425
QY 421 CTTCTGTGAGGACCGCAAGGCTGAGCCGAGGGGCTGCGGCGCTGCACCGCGCGGGT 480
Db 426 CTTCTGTGAGGACCGCAAGGCTGAGCCGAGGGGCTGCGGCGCTGCACCGCGCGGGT 485
QY 481 GCAATAGCCATCATGACCTTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAA 540
Db 486 GCAATAGCCATCATGACCTTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAA 545
QY 541 CCATGAAGAACTTTCAAAGCCTGGGAGGGCTGCATGAAATTCAGTTTCGTCTCTCCAG 600
Db 546 CCATGAAGAACTTTCAAAGCCTGGGAGGGCTGCATGAAATTCAGTTTCGTCTCTCCAG 605
QY 601 ACAGCTTCGGCGCATCTTTTGGCCCTGTATGAGTTGATGACTTACGAGACGCATT 657
Db 606 ACAGCTTCGGCGCATCTTTTGGCCCTGTATGAGTTGATGACTTACGAGACGCATT 662

RESULT 9
LOCUS BF975096 60245679F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4336722 5',
DEFINITION mRNA sequence.
```

```
ACCESSION BF975096
VERSION BF975096.1 GI:12342311
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 693)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs@mail.nih.gov
Tissue procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCML209 row: n column: 19
High quality sequence stop: 692.
Location/Qualifiers
1. .693
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4336722"
/clone_lib="NIH_MGC_48"
/tissue_type="primary B-cells from tonsils (cell line)"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: B-cells; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
BASE COUNT 179 a 167 c 162 g 185 t
ORIGIN

Query Match 21.0%; Score 592; DB 12; Length 69;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 592; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 304 CTACCGCGTCACCTGGTTACCTTCCTGGAGCCCTGCTAGGACTGTGCCGACATGTGGC 363
Db 1 CTACCGCGTCACCTGGTTACCTTCCTGGAGCCCTGCTAGGACTGTGCCGACATGTGGC 60
QY 364 CGACTTTCTCGGAGGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCTCTACTT 423
Db 61 CGACTTTCTCGGAGGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCTCTACTT 120
QY 424 CTGTGAGGACCGCAAGGCTGAGCCGAGGGGCTGCGGCGCTGCACCGCGCGGGTGCA 483
Db 121 CTGTGAGGACCGCAAGGCTGAGCCGAGGGGCTGCGGCGCTGCACCGCGCGGGTGCA 180
QY 484 AATAGGCATCATGACCTTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACCA 543
Db 181 AATAGGCATCATGACCTTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACCA 240
QY 544 TGAAGAACTTTCAAAGCCTGGGAGGGCTGCATGAAATTCAGTTTCGTCTCTCCAGACA 603
Db 241 TGAAGAACTTTCAAAGCCTGGGAGGGCTGCATGAAATTCAGTTTCGTCTCTCCAGACA 300
QY 604 GCTTCGGCGCATCTTTTGGCCCTGTATGAGTTGATGACTTACGAGACGCATTTCGTAC 663
Db 301 GCTTCGGCGCATCTTTTGGCCCTGTATGAGTTGATGACTTACGAGACGCATTTCGTAC 360
QY 664 TTTGGGACTTTGATAGCAACTTCCAGGAATGTCACACACGATGAAATATCTCTGCTGAAG 723
Db 361 TTTGGGACTTTGATAGCAACTTCCAGGAATGTCACACACGATGAAATATCTCTGCTGAAG 420
```



```

QY 724 ACAGTGGATATAAAGACAGCTTCAAGTCTTCTGTTTTTATTTCTTCAACTCTCACTTT 783
    |||||||
Db 421 ACAGTGGATATAAAGACAGCTTCAAGTCTTCTGTTTTTATTTCTTCAACTCTCACTTT 480
QY 784 CTTAGAGTTTACAGAAAAAATATTTATATACAGCTCTTTAAAAAGATCTATGCTTTGAAA 843
    |||||||
Db 481 CTTAGAGTTTACAGAAAAAATATTTATATACAGCTCTTTAAAAAGATCTATGCTTTGAAA 540
QY 844 ATAGAGAAGGAACACAGGCTCGCCAGGAGCTGCTGCAATTGGTGCGAGTTT 895
    |||||||
Db 541 ATAGAGAAGGAACACAGGCTCGCCAGGAGCTGCTGCAATTGGTGCGAGTTT 592

RESULT 10
LOCUS BG341546 670 bp mRNA linear EST 27-FEB-2001
DEFINITION 602463652F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4576548 5',
    mRNA sequence.
ACCESSION BG341546
VERSION BG341546.1 GI:13147984
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NIH-MGC http://mgc.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1289 row: g column: 13
High quality sequence stop: 650.
Location/Qualifiers
1. .670
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4576548"
/tissue_type="primary B-cells from tonsils (cell line)"
/lab_host="DH10B (phage-resistant)"
/note="Organ: B-cells; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT 200 a 122 c 134 g 214 t
ORIGIN
Query Match 19.8%; Score 558; DB 12; Length 670;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 558; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 665 TTGGGACTTTGATAGCAACTCCAGGAATGTCACACACATGAATATCTCTGCTGAAGA 724
    |||||||
Db 32 TTGGGACTTTGATAGCAACTCCAGGAATGTCACACACATGAATATCTCTGCTGAAGA 91
QY 725 CAGTGGATAAAAAACAGTCCCTTCAAGCTCTCTGTTTTTATTTCTTCAACTCTCACTTTC 784
    |||||||
Db 92 CAGTGGATAAAAAACAGTCCCTTCAAGCTCTCTGTTTTTATTTCTTCAACTCTCACTTTC 151

```

```

QY 785 TTAGAGTTTACAGAAAAAATATTTATATACGACTCTTTAAAAAGATCTATGCTTTGAAA 844
    |||||||
Db 152 TTAGAGTTTACAGAAAAAATATTTATATACGACTCTTTAAAAAGATCTATGCTTTGAAA 211
QY 845 TAGAGAAAGGAACACAGGCTCGCCAGGAGCTGCTGCAATTGGTGCGAGTTTGAATGCAA 904
    |||||||
Db 212 TAGAGAAAGGAACACAGGCTCGCCAGGAGCTGCTGCAATTGGTGCGAGTTTGAATGCAA 271
QY 905 CATTCTCCCTTACTGGGAATAACAGAACTGCAGGACCTGGGAGCATCTCTAAAGTGTCAAC 964
    |||||||
Db 272 CATTCTCCCTTACTGGGAATAACAGAACTGCAGGACCTGGGAGCATCTCTAAAGTGTCAAC 331
QY 965 GTTTTCTTATGACTTTTAGGTAGGATGAGACAGAGGTAGATCATCTCTAAAGAGCATGGTGA 1024
    |||||||
Db 332 GTTTTCTTATGACTTTTAGGTAGGATGAGACAGAGGTAGATCATCTCTAAAGAGCATGGTGA 391
QY 1025 GAGGATCAAAATGTTTTTATATCAACATCCTTTATTTATTTGATTTCATTTCAGTTTACAGTG 1084
    |||||||
Db 392 GAGGATCAAAATGTTTTTATATCAACATCCTTTATTTATTTGATTTCATTTCAGTTTACAGTG 451
QY 1085 GTGTTAGTGATAGATTTTCTTATTTCCCTTGCAGCTTTTACTTTCAAGTAACACAAC 1144
    |||||||
Db 452 GTGTTAGTGATAGATTTTCTTATTTCCCTTGCAGCTTTTACTTTCAAGTAACACAAC 511
QY 1145 TCTTCCATCAGCCCATGATCTATAGGACCTCTCTAATGAGAGTATCTGGGTGATTGTGACC 1204
    |||||||
Db 512 TCTTCCATCAGCCCATGATCTATAGGACCTCTCTAATGAGAGTATCTGGGTGATTGTGACC 571
QY 1205 CCAACCATCTCTCCAAA 1222
    |||||||
Db 572 CCAACCATCTCTCCAAA 589

RESULT 11
LOCUS BF238155
DEFINITION 601811880F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4054915 5',
    mRNA sequence.
ACCESSION BF238155
VERSION BF238155.1 GI:11152074
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NIH-MGC http://mgc.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM895 row: p column: 20
High quality sequence stop: 541.
Location/Qualifiers
1. .541
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4054915"
/tissue_type="primary B-cells from tonsils (cell line)"
/lab_host="DH10B (phage-resistant)"
/note="Organ: B-cells; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT 200 a 122 c 134 g 214 t
ORIGIN
Query Match 19.8%; Score 558; DB 12; Length 670;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 558; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 665 TTGGGACTTTGATAGCAACTCCAGGAATGTCACACACATGAATATCTCTGCTGAAGA 724
    |||||||
Db 32 TTGGGACTTTGATAGCAACTCCAGGAATGTCACACACATGAATATCTCTGCTGAAGA 91
QY 725 CAGTGGATAAAAAACAGTCCCTTCAAGCTCTCTGTTTTTATTTCTTCAACTCTCACTTTC 784
    |||||||
Db 92 CAGTGGATAAAAAACAGTCCCTTCAAGCTCTCTGTTTTTATTTCTTCAACTCTCACTTTC 151

```

Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

BASE COUNT	120 a	150 c	142 g	129 t	
ORIGIN					
Query Match	19.2%	Score 540;	DB 12;	Length 541;	
Best Local Similarity	100.0%;	Pred. No. 0;	Mismatches 0;	Indels 0;	Gaps 0;
Matches 540;	Conservative	0;			
Qy	4	GAACCATCATTAATTGAAGTGAAGATTTTCTGGCTGAGACTTGCGAGGAGGCAAGAAGA	63		
Db	2	GAACCATCATTAATTGAAGTGAAGATTTTCTGGCTGAGACTTGCGAGGAGGCAAGAAGA	61		
Qy	64	CACCTCTGGACACCATATGGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATT	123		
Db	62	CACCTCTGGACACCATATGGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATT	121		
Qy	124	CAAAAATGTCCGCTGGGCTAAGGTCGGCTGAGACCTACCTGTGCTACGTAGTGAAGAG	183		
Db	122	CAAAAATGTCCGCTGGGCTAAGGTCGGCTGAGACCTACCTGTGCTAGTGAAGAG	181		
Qy	184	CGGTGACAGTGTACATCCTTTTCACTGGACTTTGTTATCTTCGCAATAAAGACGGGTG	243		
Db	182	CGGTGACAGTGTACATCCTTTTCACTGGACTTTGTTATCTTCGCAATAAAGACGGGTG	241		
Qy	244	CCACCTGGAATTCCTCCCTCCGTACATCTCGACTGGGACCTAGACCTGGCGGCTG	303		
Db	242	CCACCTGGAATTCCTCCCTCCGTACATCTCGACTGGGACCTAGACCTGGCGGCTG	301		
Qy	304	CTACCGCGTCACTGTTTCACTCTGGAGCCCTGCTACGACTGTGCCGACATGTGCG	363		
Db	302	CTACCGCGTCACTGTTTCACTCTGGAGCCCTGCTACGACTGTGCCGACATGTGCG	361		
Qy	364	CGACTTTCTGCGAGGAGACCCCAACCTCAGTCTGAGAGTCTTCACCGCGGCTCTACTT	423		
Db	362	CGACTTTCTGCGAGGAGACCCCAACCTCAGTCTGAGAGTCTTCACCGCGGCTCTACTT	421		
Qy	424	CTGTGAGGACCCGAGGCTGAGCCGAGGCGGCTGCGGCGCTGACCGCGCGGGTGCA	483		
Db	422	CTGTGAGGACCCGAGGCTGAGCCGAGGCGGCTGCGGCGCTGACCGCGCGGGTGCA	481		
Qy	484	AATAGCCATCATGACCTTCAAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACA	543		
Db	482	AATAGCCATCATGACCTTCAAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACA	541		

RESULT 12  
AW504807  
LOCUS  
DEFINITION  
UI-HF-BN0-aln-f-05-0-UI.r1 NIH\_MGC\_50 Homo sapiens cDNA clone  
IMAGE:3080216 5', mRNA sequence.  
ACCESSION  
AW504807  
VERSION  
AW504807.1 GI:7142474  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
NIH-MGC http://mgc.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-r@mail.nih.gov  
Eco RI site shown at the beginning of the sequence.  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
cDNA Library Preparation: M.B. Soares Lab  
cDNA Library Arrayed by: M.B. Soares Lab  
DNA Sequencing by: M.B. Soares Lab  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:

www.bio.llnl.gov/bbrp/image/image.html  
Seq primer: M13 Forward.  
FEATURES  
source  
1..570  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3080216"  
/clone\_lib="NIH\_MGC\_50"  
/tissue\_type="lymph"  
/cell\_type="germinal center B cells"  
/cell\_line="MGC85"  
/lab\_host="DH10B (LTI)"  
/note="Vector: pT7T3-Pac; Site\_1: NotI; Site\_2: Eco RI;  
Constructed from size fractionated cytoplasmic mRNA  
(3.5-4.4kb). Directionally cloned. Cells provided by  
Louis M. Staudt, Ph.D. Library preparation by Maria de  
Fatima Bonaldo, Ph.D. and M. Bento Soares, Ph.D. "

BASE COUNT	178 a	106 c	120 g	165 t	1 others
ORIGIN					
Query Match	18.6%	Score 523;	DB 10;	Length 570;	
Best Local Similarity	100.0%;	Pred. No. 0;	Mismatches 0;	Indels 0;	Gaps 0;
Matches 523;	Conservative	0;			
Qy	1111	TTTCCCTTGACGCTTTACTTTCAAGTAACACAAACTCTTCCATCAGGCCATGATCATAGG	1170		
Db	8	TTTCCCTTGACGCTTTACTTTCAAGTAACACAAACTCTTCCATCAGGCCATGATCATAGG	67		
Qy	1171	ACCTCCTAATGAGAGATATCTGGTGATTTGTGACCCCAACCACTCTCCAAAGCATTAAT	1230		
Db	68	ACCTCCTAATGAGAGATATCTGGTGATTTGTGACCCCAACCACTCTCCAAAGCATTAAT	127		
Qy	1231	ATCCCAATCATGCGCTGTATGTTTAAATCAGCAAGAAGCATGTTTATGTTCTACAAAAG	1290		
Db	128	ATCCCAATCATGCGCTGTATGTTTAAATCAGCAAGAAGCATGTTTATGTTCTACAAAAG	187		
Qy	1291	AAGATTGTTATGGTGGGATGGAGGTATAGACCATGATGTCACCTTCAAGCTACTTTT	1350		
Db	188	AAGATTGTTATGGTGGGATGGAGGTATAGACCATGATGTCACCTTCAAGCTACTTTT	247		
Qy	1351	ATAAAGGATCTTAAATGGGAGGAGGACTGTGAACAGACACACCTATATATGGTTCA	1410		
Db	248	ATAAAGGATCTTAAATGGGAGGAGGACTGTGAACAGACACACCTATATATGGTTCA	307		
Qy	1411	TGCTGGAAGTAGCAAAATCTTCTGGAACGCAAACTCTTTAAGGAAGTCTCTAAATTTAGA	1470		
Db	308	TGCTGGAAGTAGCAAAATCTTCTGGAACGCAAACTCTTTAAGGAAGTCTCTAAATTTAGA	367		
Qy	1471	AACACCCCAAACTTCCACATATCATATTTAGCAACAATTTGAAGGAAGTTGCTTGAATG	1530		
Db	368	AACACCCCAAACTTCCACATATCATATTTAGCAACAATTTGAAGGAAGTTGCTTGAATG	427		
Qy	1531	TTGGGGAGAGGAAATCTATTGGCTCTCTGGGTCTCTTCATCTCAGAAATGCCAATCAG	1590		
Db	428	TTGGGGAGAGGAAATCTATTGGCTCTCTGGGTCTCTTCATCTCAGAAATGCCAATCAG	487		
Qy	1591	GTCAAGTTTGTGCTACATTTTGTATGTTGTGATGCTTCTCTCCCA	1633		
Db	488	GTCAAGTTTGTGCTACATTTTGTATGTTGTGATGCTTCTCTCCCA	530		

RESULT 13  
BG757392  
LOCUS  
DEFINITION  
mRNA sequence.  
602711022F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4851580 5',  
BG757392  
ACCESSION  
BG757392.1 GI:14068045  
VERSION  
BG757392.1  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

[illegible]

Db 69 TTCTTTACCAATTCAAAATGTCCGCTGGGCTAAGGGTCGGCTGAGACCTACCTGTGCT 128  
QY 171 ACGTAGTGAAGGGCTGACAGTGTACATCCTTTTCACTGACACTTTGGTATCTTCGCA 230  
Db 129 ACGTAGTGAAGGGCTGACAGTGTACATCCTTTTCACTGACACTTTGGTATCTTCGCA 188  
QY 231 ATAAGAAGGGTGCACCGTGGAAATGTCTTCTCCGCTACATCTCGGACTGGGACCTAG 290  
Db 189 ATAAGAAGGGTGCACCGTGGAAATGTCTTCTCCGCTACATCTCGGACTGGGACCTAG 248  
QY 291 ACCCTGGCGGTGCTACCGCTGACCTGCTGCTACCTGCTGAGCCCTGCTGACACTGTG 350  
Db 249 ACCCTGGCGGTGCTACCGCTGACCTGCTGCTGCTGAGCCCTGCTGACACTGTG 308  
QY 351 CCCGACATGTGCGGACCTTCTGCGAGGGAACCCCAACCTCAGTCTGAGGATCTTCACCG 410  
Db 309 CCCGACATGTGCGGACCTTCTGCGAGGGAACCCCAACCTCAGTCTGAGGATCTTCACCG 368  
QY 411 CGCGCTCTACTTCTGTGAGGACCGCAAGGCTGAGCCGAGGGCTGCGGGGGCTGCACC 470  
Db 369 CGCGCTCTACTTCTGTGAGGACCGCAAGGCTGAGCCGAGGGCTGCGGGGGCTGCACC 428  
QY 471 GCGCGGGGTGCMAATAGCCATCATGACCTTCARAG 506  
Db 429 GCGCGGGGTGCMAATAGCCATCATGACCTTCARAG 464

RESULT 15  
AW978582/c  
LOCUS AW978582 604 bp mRNA linear EST 02-JUN-2000  
DEFINITION AW978582 MAGP resequences, MAGP Homo sapiens cDNA, mRNA sequence.  
ACCESSION AW978582  
VERSION AW978582.1 GI:8169850  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 604)  
AUTHORS Hegde, P., Qi, R., Abernathy, K., Dharap, S., Gaspard, R., Gay, C., Holt, I.E., Saeed, A.I., Sharov, V., Lee, N.H., Yeatman, T.J. and Quackenbush, J.  
TITLE Assessment of gene expression patterns in a model of colon tumor metastasis using a 19,200 element cDNA microarray  
JOURNAL Unpublished (2000)  
COMMENT The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 3528  
Fax: 301 838 0208  
Email: johnq@tigr.org  
Plate: 394  
Seq primer: Forward.

FEATURES  
source  
1..604  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="MAGE resequences, MAGP"  
/note="Vector: pBluescriptSKm"  
BASE COUNT 190 a 121 c 109 g 184 t  
ORIGIN

Query Match 15.9%; Score 448; DB 10; Length 604;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 448; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1197 TTGTGACCCCAACCATCTCTCCAAGCATTAATCAATCATCGCTGTATGTTTAA 1256  
Db 459 TTGTGACCCCAACCATCTCTCCAAGCATTAATCAATCATCGCTGTATGTTTAA 400  
QY 1257 TCAGCAGAGCATGTTTTATCTTTGTACAAAGAGATTGTTATGGGTGGGATGGAGG 1316  
Db 399 TCAGCAGAGCATGTTTTATCTTTGTACAAAGAGATTGTTATGGGTGGGATGGAGG 340

QY 1317 TATAGACCATGATGCTCACCTTCAAGCTACTTTAATAAGGATCTTAAAAATGGCAGGA 1376  
Db 339 TATAGACCATGATGCTCACCTTCAAGCTACTTTAATAAGGATCTTAAAAATGGCAGGA 280  
QY 1377 GGACTGTGAACAAGACACCCCTAATAATGGGTTGATGTCTGAAGTAGCAAACTTCTTGGAA 1436  
Db 279 GGACTGTGAACAAGACACCCCTAATAATGGGTTGATGTCTGAAGTAGCAAACTTCTTGGAA 220  
QY 1437 ACGCAAACTCTTTTAAGGAAGTCCCTAATTTAGAAACACCCCAAACTTCACATATCATA 1496  
Db 219 ACGCAAACTCTTTTAAGGAAGTCCCTAATTTAGAAACACCCCAAACTTCACATATCATA 160  
QY 1497 ATTACCAACAATTCGAAGGAAGTTGCTTGAATGTTGGGAGAGAAATCTATTGGGCTC 1556  
Db 159 ATTACCAACAATTCGAAGGAAGTTGCTTGAATGTTGGGAGAGAAATCTATTGGGCTC 100  
QY 1557 TCGTGGGTCTCTTCATCTCAGAAATGCCAATCAGGTCAAGGTTTCTACATTTGTATGT 1616  
Db 99 TCGTGGGTCTCTTCATCTCAGAAATGCCAATCAGGTCAAGGTTTCTACATTTGTATGT 40  
QY 1617 GTGTGATGCTTCTCCCAAGGTATATTA 1644  
Db 39 GTGTGATGCTTCTCCCAAGGTATATTA 12

RESULT 16  
BF975166  
LOCUS BF975166 942 bp mRNA linear EST 22-JAN-2001  
DEFINITION BF975166 Homo sapiens cDNA clone IMAGE:433639 5', mRNA sequence.  
ACCESSION BF975166  
VERSION BF975166.1 GI:12342381  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 942)  
AUTHORS NIH-MGC <http://mcc.nci.nih.gov/>  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLCM1207 row: a column: 16  
High quality sequence stop: 707.

FEATURES  
source  
1..942  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:433639"  
/clone\_lib="NIH\_MGC\_48"  
/tissue\_type="primary B-cells from tonsils (cell line)"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: B-cells; Vector: pOTB7; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."  
BASE COUNT 249 a 209 c 251 g 233 t  
ORIGIN

Query Match 15.9%; Score 448; DB 12; Length 942;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 498; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 122 TTCAAAATGTCGCTGGCTAAGGTCGGGTGAGACTACCTGTGCTAGCTAGTGAAG 181  
|||||  
DB 121 TTCAAAATGTCGCTGGCTAAGGTCGGGTGAGACTACCTGTGCTAGCTAGTGAAG 180  
|||||  
QY 182 AGGCGTGACAGTGCTACATCTTTTCACTGGACTTTGTTTCTTCGCAATAAGAACGGC 241  
|||||  
DB 181 AGGCGTGACAGTGCTACATCTTTTCACTGGACTTTGTTTCTTCGCAATAAGAACGGC 240  
|||||  
QY 242 TGCCACGTGGAAATGCTCTTCCTCGCTACATCTCGGACTGGGACTAGACCTGGCGGC 301  
|||||  
DB 241 TGCCACGTGGAAATGCTCTTCCTCGCTACATCTCGGACTGGGACTAGACCTGGCGGC 300  
|||||  
QY 302 TGCTACCGGTCACCTGTTTCACTCTCGGAGCCCTGCTAGGACTGTGCGGACATGTG 361  
|||||  
DB 301 TGCTACCGGTCACCTGTTTCACTCTCGGAGCCCTGCTAGGACTGTGCGGACATGTG 360  
|||||  
QY 362 GCGGACTTTCTCGAGGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCTCTAC 421  
|||||  
DB 361 GCGGACTTTCTCGAGGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCTCTAC 420  
|||||  
QY 422 TTCTGTGAGGACCGCAAGGCTGAGCCCGAGGGCTGCGGCGGCTGCACCGCGCGGGTG 481  
|||||  
DB 421 TTCTGTGAGGACCGCAAGGCTGAGCCCGAGGGCTGCGGCGGCTGCACCGCGCGGGTG 480  
|||||  
QY 482 CAATAGGACATCATGACCTTCAAGATATTTTACTGCTGGAATCTTTGTAGAAAAC 541  
|||||  
DB 481 CAATAGGACATCATGACCTTCAAGATATTTTACTGCTGGAATCTTTGTAGAAAAC 540  
|||||  
QY 542 CATGAAAGAACCTTTCAAGCCCTGGGAAGGCTGCATGAAATTCAGTTCGCTCTCCAGA 601  
|||||  
DB 541 CATGAAAGAACCTTTCAAGCCCTGGGAAGGCTGCATGAAATTCAGTTCGCTCTCCAGA 600  
|||||  
QY 602 CAGCTTCGGCGCATCCTTT 620  
|||||  
DB 601 CAGCTTCGGCGCATCCTTT 619  
|||||

RESULT 17  
AI016902/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
FEATURES  
source

AI016902 442 bp mRNA linear EST 17-MAR-1999  
ou31d03.x1 Soares\_NFL\_T\_GBC\_S1 Homo sapiens cDNA clone  
IMAGE:1627877 3', mRNA sequence.

AI016902  
EST.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 442)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 426 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 433.  
Location/Qualifiers  
1. .442  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1627877"  
/clone\_lib="Soares\_NFL\_T\_GBC\_S1"  
/lab\_host="DH10B"  
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with

a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
Equal amounts of plasmid DNA from three normalized  
libraries (fetal lung NBHL19W, testis NHT, and B-cell  
NCI-CGAP-GCB1) were mixed, and ss circles were made in  
vitro. Following HAP purification, this DNA was used as  
tracer in a subtractive hybridization reaction. The driver  
was PCR-amplified cDNAs from pools of 5,000 clones made  
from the same 3 libraries. The pools consisted of  
I.M.A.G.E. clones 297480-302087, 682632-687239,  
726408-728711, and 729096-731399. Subtraction by Bento  
Soares and M. Fatima Bonaldo.

BASE COUNT 166 a 66 c 69 g 141 t  
ORIGIN

Query Match 15.7%; Score 442; DB 9; Length 442;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 442; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2211 AATGTAAGGAGCCAAATAAGGATCCCTATTTGCTCTTTTGGTGTCTATTTGTCCTTAAC 2270  
|||||  
DB 442 AATGTAAGGAGCCAAATAAGGATCCCTATTTGCTCTTTTGGTGTCTATTTGTCCTTAAC 383  
|||||  
QY 2271 AACTGTCTTTGACAGTGAGAAAATATTCAGAAATACCATATCCCTGTGCGGTATATACC 2330  
|||||  
DB 382 AACTGTCTTTGACAGTGAGAAAATATTCAGAAATACCATATCCCTGTGCGGTATATACC 323  
|||||  
QY 2331 TAGCAACCTTTGCAATGAAGATGAGCAGATCCACAGGAAACTTCAATGCCACAACCTGTCT 2390  
|||||  
DB 332 TAGCAACCTTTGCAATGAAGATGAGCAGATCCACAGGAAACTTCAATGCCACAACCTGTCT 263  
|||||  
QY 2391 TATTTAATCTTATTTGATCAATAGTTTGTAAAAGAGTTTAAAATTTGTTACTTTCATGTAT 2450  
|||||  
DB 262 TATTTAATCTTATTTGATCAATAGTTTGTAAAAGAGTTTAAAATTTGTTACTTTCATGTAT 203  
|||||  
QY 2451 CATTTATATTTATATTTTGGCTCTAATGATTTTATTAACATGATTTCCCTTTCT 2510  
|||||  
DB 202 CATTTATATTTATATTTTGGCTCTAATGATTTTATTAACATGATTTCCCTTTCT 143  
|||||  
QY 2511 GATATATTGAATGGAGTCTCAAGCTTCAATAATTTATACTTTAGAAATGATTCATAT 2570  
|||||  
DB 142 GATATATTGAATGGAGTCTCAAGCTTCAATAATTTATACTTTAGAAATGATTCATAT 83  
|||||  
QY 2571 AACACGATGTAATTTAATGTAACATTCAGTAATGTTGCTACGAGCCATTTCTCTTTGATTT 2630  
|||||  
DB 82 AACACGATGTAATTTAATGTAACATTCAGTAATGTTGCTACGAGCCATTTCTCTTTGATTT 23  
|||||  
QY 2631 TTAGTAAACTTTTATGACAGCA 2652  
|||||  
DB 22 TTAGTAAACTTTTATGACAGCA 1  
|||||

RESULT 18  
BG170756  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
FEATURES  
source

BG170756 511 bp mRNA linear EST 06-FEB-2001  
602323871F1 NIH\_MGC\_89 Homo sapiens cDNA clone IMAGE:4426960 5',  
mRNA sequence.

BG170756  
EST.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 511)  
NIH-MGC http://mgc.ncbi.nlm.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: ATCC  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Plate: LAM10174 row: n column: 17  
 High quality sequence stop: 511.  
 Location/Qualifiers

FEATURES

1..511  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4426960"  
 /clone\_lib="NIH\_MGC\_89"  
 /tissue\_type="hyponephroma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: kidney; Vector: PCMV-SPORT6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.3 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH\_MGC Library."

BASE COUNT 181 a 91 c 114 g 125 t  
 ORIGIN

Query Match 13.5%; Score 380; DB 12; Length 511;  
 Best Local Similarity 99.8%; Pred. No. 0;  
 Matches 500; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1318 ATAGACCATGCTGCTCAAGCTTCAAGTCTTAAATAAGGATCTTAAATGGGAGGAG 1377

Db 1 ATAGACCATGCTGCTCAAGCTTCAAGTCTTAAATAAGGATCTTAAATGGGAGGAG 60

QY 1378 GACTGTGACAGACACCCCTAATATAGTGGTGTGATCTGAGTAGCAAACTCTTCGGAAA 1437

Db 61 GACTGTGACAGACACCCCTAATATAGTGGTGTGATCTGAGTAGCAAACTCTTCGGAAA 120

QY 1438 CGCAAACTCTTTAAGGAAGTCCCTAATTTAGAAACACCCACAACTTCACATATCAATA 1497

Db 121 CGCAAACTCTTTAAGGAAGTCCCTAATTTAGAAACACCCACAACTTCACATATCAATA 180

QY 1498 TTAGCAAACTTTGGAAGGAAGTCTTGAATGTTGGGAGAGGAAAATCTATTGGCTCT 1557

Db 181 TTAGCAAACTTTGGAAGGAAGTCTTGAATGTTGGGAGAGGAAAATCTATTGGCTCT 240

QY 1558 CTTGGCTCTCTTCATCTCAGAAATGCCAATCAGGTCAGAGTTTGCTACATTTGTATGTG 1617

Db 241 CTTGGCTCTCTTCATCTCAGAAATGCCAATCAGGTCAGAGTTTGCTACATTTGTATGTG 299

QY 1618 TGTGATGCTCTCCCAAGGTATATTAACATATATAAGAGAGTTGTGACAAACAGAAATGA 1677

Db 300 TGTGATGCTCTCCCAAGGTATATTAACATATATAAGAGAGTTGTGACAAACAGAAATGA 359

QY 1678 TAAAGCTGCGAACCGTGGCACACGCTCATAGTTCTAGCTGCTGGAGGTTGAGGAGGGA 1737

Db 360 TAAAGCTGCGAACCGTGGCACACGCTCATAGTTCTAGCTGCTGGAGGTTGAGGAGGGA 419

QY 1738 GGATGGCTTGACACAGGTGTTCAAGGCCAGCTGGGCAACATACAGATCCCTGCTCT 1797

Db 420 GGATGGCTTGACACAGGTGTTCAAGGCCAGCTGGGCAACATACAGATCCCTGCTCT 479

QY 1798 CAAAAAATAAAAAAAAAAAAAA 1818

Db 480 CAAAAAATAAAAAAAAAAAAAA 500

RESULT 19

AW401901

LOCUS

DEFINITION

UI-HF-BK0-aam-c-04-0-UI-rl NIH\_MGC\_36 Homo sapiens cDNA clone

IMAGE:3054318 5', mRNA sequence.

ACCESSION

AW401901

VERSION

AW401901.1 GI:6920587

KEYWORDS

human.

SOURCE

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 374)  
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
 Eco RI site shown at the beginning of the sequence.  
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
 cDNA Library Preparation: M.B. Soares Lab  
 cDNA Library Arrayed by: M.B. Soares Lab  
 DNA Sequencing by: M.B. Soares Lab  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://bio.llnl.gov/bbrp/image/image.html>  
 Seq primer: M13 Forward

FEATURES  
 source  
 1..374  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3054318"  
 /clone\_lib="NIH\_MGC\_36"  
 /tissue\_type="lymph"  
 /cell\_type="germinal center B cells"  
 /cell\_line="MGC85"  
 /lab\_host="DH10B (LT1)"  
 /note="Vector: pT7T3-Pac; Site\_1: NotI; Site\_2: Eco RI;  
 Constructed from size fractionated cytoplasmic mRNA  
 (0.5-1.5kb). Directionally cloned. Cells provided by Louis  
 M. Staudt, Ph.D. Library preparation by Maria de Fatima  
 Bonaldo, Ph.D. and M. Bento Soares, Ph.D."

BASE COUNT 112 a 70 c 77 g 115 t  
 ORIGIN

Query Match 13.1%; Score 369; DB 10; Length 374;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 369; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1084 GGTGTAGTGTAGATATTTCTTATTTCTTTCCCTTGACGTTTACTTTCAAGTAACACAAA 1143

Db 6 GGTGTAGTGTAGATATTTCTTATTTCTTTCCCTTGACGTTTACTTTCAAGTAACACAAA 65

QY 1144 CTCCTTCATCAGGCCATGATCTATAGGACCTCCTAATCAGAGATATCTGGTGTGTCAC 1203

Db 66 CTCCTTCATCAGGCCATGATCTATAGGACCTCCTAATCAGAGATATCTGGTGTGTCAC 125

QY 1204 CCCAACCATCTCTCCAAAGCATTAATATCAATCATCGGCTGTATGTTTAAATCAGCAG 1263

Db 126 CCCAACCATCTCTCCAAAGCATTAATATCAATCATCGGCTGTATGTTTAAATCAGCAG 185

QY 1264 AAGCATGTTTTATGTTTTGTACAAAAGAGATTGTTATGGTGGGATGGAGGTATAGAC 1323

Db 186 AAGCATGTTTTATGTTTTGTACAAAAGAGATTGTTATGGTGGGATGGAGGTATAGAC 245

QY 1324 CATGCATGGTCACCTTCAAGCTACTTTAATAAGGATCTTAAATGGCAGCAGGACTGT 1383

Db 246 CATGCATGGTCACCTTCAAGCTACTTTAATAAGGATCTTAAATGGCAGCAGGACTGT 305

QY 1384 GAACAGACACCCCTAATAATGGTGTGATGTCGTAAGTAGCAAACTTCTGAAACGCCAAA 1443

Db 306 GAACAGACACCCCTAATAATGGTGTGATGTCGTAAGTAGCAAACTTCTGAAACGCCAAA 365

QY 1444 CTCCTTTTAA 1452

Db 366 CTCCTTTTAA 374

RESULT 20

AA831307/c

LOCUS

DEFINITION

oc70a06.s1 NCL\_CGAP\_GC81 Homo sapiens cDNA clone IMAGE:1355026 3'

similar to contains Alu repetitive element; mRNA sequence.

ACCESSION

AA831307

LOCUS

AA831307

DEFINITION

353 bp mRNA linear EST 20-FEB-1998

oc70a06.s1 NCL\_CGAP\_GC81 Homo sapiens cDNA clone IMAGE:1355026 3'

similar to contains Alu repetitive element; mRNA sequence.

ACCESSION

AA831307

LOCUS

AA831307

DEFINITION

353 bp mRNA linear EST 20-FEB-1998

oc70a06.s1 NCL\_CGAP\_GC81 Homo sapiens cDNA clone IMAGE:1355026 3'

similar to contains Alu repetitive element; mRNA sequence.

ACCESSION

AA831307

AA831307.1	GI:2904406	Db		53	CGCAACCGTGGCACACGCTCATAGTTCCTAGCTGCTGGAGGTTGAGGAGGGA	1
VERSION	EST.	RESULT 21				
KEYWORDS	human.	AW135547/c				
ORGANISM	Homo sapiens	LOCUS				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	DEFINITION				
AUTHORS	1 (bases 1 to 353)	NCI-CGAP				
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.	IMAGE:2714208				
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index	AW135547				
COMMENT	Unpublished (1997)	AW135547.1				
	Contact: Robert Strausberg, Ph.D.	GI:6139680				
	Email: cgabbs@mail.nih.gov	EST.				
	Tissue Procurement: Louis M. Staudt, M.D., David Allman, Ph.D., Gerald Marti, M.D.	Homo sapiens				
	cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
	CDNA Library Arrayed by: Greg Lennon, Ph.D.	1 (bases 1 to 476)				
	DNA Sequencing by: Washington University Genome Sequencing Center	NCI-CGAP				
	clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:	Unpublished (1997)				
	www-bio.llnl.gov/bbrp/image/image.html	Contact: Robert Strausberg, Ph.D.				
	Seq primer: -40m13 fwd. ET from Amersham	Email: cgabbs@mail.nih.gov				
	High quality sequence stop: 304.	Oligo-dr track not found, Not I site shown in beginning of sequence				
	Location/Qualifiers	is likely internal to the message. cDNA Library Preparation: M.B.				
FEATURES	1..353	Soares Lab Clone distribution: NCI-CGAP clone distribution				
source	/organism="Homo sapiens"	information can be found through the I.M.A.G.E. Consortium/LLNL at:				
	/db_xref="taxon:9606"	www-bio.llnl.gov/bbrp/image/image.html				
	/clone="IMAGE:1355026"	Seq primer: M13 Forward				
	/clone_lib="NCI-CGAP_GCB1"	POLYA=No.				
	/tissue_type="germinal center B cell"	Location/Qualifiers				
	/lab_host="DH10B"	1..476				
	/note="Vector: p7T3D-Pac (Pharmacia) with a modified	/organism="Homo sapiens"				
	polylinker. Site.1: Not I; Site.2: Eco RI; 1st strand cDNA	/db_xref="taxon:9606"				
	was prepared from human tonsillar cells enriched for	/clone="IMAGE:2714208"				
	germinal center B cells by flow sorting (CD20+, IgD-),	/clone_lib="NCI-CGAP_Sub3"				
	provided by Dr. Louis M. Staudt (NCI), Dr. David Allman	/lab_host="DH10B (Life Technologies)"				
	(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was	/note="Vector: p7T3D-Pac (Pharmacia) with a modified				
	primed with a Not I - oligo(dT) primer	polylinker. Site.1: Not I; Site.2: Eco RI; The				
	5'-TGTTACCAATCTGAAGGCGGCGGCTCATTTTCTTTT-3'	NCI-CGAP Sub3 library is a subtracted library derived from				
	]. Double-stranded cDNA was ligated to Eco RI adaptors	the NCI-CGAP Sub1 library, which is a subtracted library				
	(Pharmacia), digested with Not I and cloned into the Not I	derived from BI. BI constitutes a mixture of 21				
	and Eco RI sites of the modified p7T3 vector. Library	normalized or subtracted NCI-CGAP libraries: NCI-CGAP_Co4				
	went through one round of normalization, and was	, NCI-CGAP_Pr22, NCI-CGAP_Pr28, NCI-CGAP_Col10,				
	constructed by Bento Soares and M. Fatima Bonaldo."	NCI-CGAP_Col16, NCI-CGAP_Kid5, NCI-CGAP_Kid12,				
BASE COUNT	96 a 81 c 63 g 113 t	NCI-CGAP_Kid3, NCI-CGAP_Kid11, NCI-CGAP_Lym2,				
ORIGIN		NCI-CGAP_Pr2, NCI-CGAP_Co8, NCI-CGAP_Col1, NCI-CGAP_Le12,				
		NCI-CGAP_Brn23, NCI-CGAP_Lu5, NCI-CGAP_Lu24,				
		NCI-CGAP_Lu19, NCI-CGAP_GC4, NCI-CGAP_GG6,				
		NCI-CGAP_Brn25. These 21 libraries were pooled and a				
		single-stranded DNA preparation of the resulting mixture				
		was used as a tracer in a subtractive hybridization with				
		a driver whose composition is detailed below:				
		NCI-CGAP_Kid3 pool 1 LLAM 3334-3337, 3682-3683,				
		3798-3803 (IMAGE Clonides 1322376-1323911, 1456008-1456775				
		, 1500552-1502855); NCI-CGAP_Kid5 pool 1 LLAM 3338-3342				
		, 3722-3725, 3776-3778 (IMAGE Clonides 1323912-1325831,				
		1471368-1472903, 1492104-1493255); NCI-CGAP_Lu5 pool 1				
		LLAM 3575-3582, 3851-3854 (IMAGE Clonides 1414920-1417991,				
		1520904-1522439); NCI-CGAP_GC4 pool 1 LLAM 3164-3167,				
		3716-3720, 3733-3735 (IMAGE Clonides 1257096-1258631,				
		1460664-1470983, 1475592-1476743); NCI-CGAP_Pr22 pool 1				
		LLAM 2457-2459, 2758-2759, 3062-3068 (IMAGE Clonides				
		985608-986759, 1101192-1101959, 1217928-1220615);				
		NCI-CGAP_Col10 pool 1 LLAM 2684-2653, 2871-2872 (IMAGE				
		Clonides 1057416-1061255, 1144584-1145351). Subtraction				
		was performed as previously described (Bonaldo, Lennon &				
		Soares (1996): Normalization and Subtraction: Two				
		Approaches To Facilitate Gene Discovery. Genome Research				
		6, 791-806.				
		TAG.LIB=NCI-CGAP_Lu5				
		TAG.TISSUE=lung				

BASE COUNT		145 a	102 c	86 g	143 t
ORIGIN					
Query Match		12.4%; Score 349; DB 10; Length 476;			
Best Local Similarity		99.8%; Pred. No. 0;			
Matches		399; Conservative	0; Mismatches	1; Indels	0; Gaps
QY	832	TATGCTCTTGAATAAGAGAGAACACAGCTCTGGCCAGGAGCTGCTGCAATTGGTGCA 891			
Db	476	TATGCTCTTGAATAAGAGAGAACACAGCTCTGGCCAGGAGCTGCTGCAATTGGTGCA 417			
QY	892	GTTTGAATGCAACATGTCCTCCCTACTGGGAATAACAGAACTGCAGGACCTGGGAGCATC 951			
Db	416	GTTTGAATGCAACATGTCCTCCCTACTGGGAATAACAGAACTGCAGGACCTGGGAGCATC 357			
QY	952	CTAAAGTGTCAACGTTTTTCTATGACTTTTAGTAGATGAGAGCAGAGAGTATGATCCTA 1011			
Db	356	CTAAAGTGTCAACGTTTTTCTATGACTTTTAGTAGAGATGAGAGCAGAGAGTATGATCCTA 297			
QY	1012	AAAAGCATGGTGAGAGGATCAAAATGTTTTATATCAACATCCCTTTATTATTTGATTCATT 1071			
Db	296	AAAAGCATGGTGAGAGGATCAAAATGTTTTATATCAACATCCCTTTATTATTTGATTCATT 237			
QY	1072	TGAGTTAAACAGTGGTGTAGTAGATAGATTTTCTATTTTCCCTTGAGCTTTACTTTTC 1131			
Db	236	TGAGTTAAACAGTGGTGTAGTAGATAGATTTTCTATTTTCCCTTGAGCTTTACTTTTC 177			
QY	1132	AAGTACACAACTCTTCCATCAGGCCATGATCTATAGGACCTCCTCAATCAGAGTATCTG 1191			
Db	176	AAGTACACAACTCTTCCATCAGGCCATGATCTATAGGACCTCCTCAATCAGAGTATCTG 117			
QY	1192	GGTGATTGTGACCCCAACCATCTCTCCAAAGCATTAATA 1231			
Db	116	GGTGATTGTGACCCCAACCATCTCTCCAAAGCATTAATA 77			
RESULT 22					
AW452648/c					
LOCUS		AW452648	464 bp	mRNA	linear
DEFINITION		UI-H-B13-alu-f-11-0-UI.sl NCI_CGAP_Sub5 Homo sapiens cDNA clone			
IMAGE		IMAGE:3068708 3', mRNA sequence.			
ACCESSION		AW452648			
VERSION		AW452648.1 GI:6993424			
KEYWORDS		EST.			
SOURCE		human.			
ORGANISM		Homo sapiens			
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS		1 (bases 1 to 464)			
TITLE		NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.			
JOURNAL		National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index			
COMMENT		Unpublished (1997)			
		Contact: Robert Strausberg, Ph.D.			
		Email: cgaps-r@mail.nih.gov			
		The sequence contained an oligo-dT track that was present in the			
		oligonucleotide that was used to prime the synthesis of first			
		strand cDNA and therefore this may represent a bonafide poly A			
		tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:			
		NCI-CGAP clone distribution information can be found through the			
		I.M.A.G.E. Consortium/LLNL at:			
		www.bio.llnl.gov/bbrp/image/image.html The following repetitive			
		elements were found in this cDNA sequence: 4-123, >ALU			
		Seq primer: M13 Forward			
		POLYA=Yes.			
FEATURES		Location/Qualifiers			
source		1..464	/organism="Homo sapiens"		
			/db_xref="taxon:9606"		
			/clone="IMAGE:3068708"		
			/clone_lib="NCI_CGAP_Sub5"		
			/lab_host="DH10B (Life Technologies)"		
			/note="Vector: pT73D-Pac (Pharmacia) with a modified		

polylinker: Site1: Not I; Site 2: Eco RI; NCI_CGAP_Sub5		is a substracted library derived from NCI_CGAP_Sub4. The			
NCI_CGAP_Sub5 library had 3 million recombinants. A		single-stranded DNA preparation of NCI_CGAP_Sub4 was used			
as a tracer in a subtractive hybridization with a driver		comprising: the IMAGE pool (NCI_CGAP_Kid3 pool 1 L1AM			
3334-3337, 3682-3683, 3798-3803 (IMAGE ClonEids		1323376-1323911, 1456008-1456775 1500552-1502855);			
NCI_CGAP_Kid5 pool 1 L1AM 3338-3342, 3722-3725, 3776-3778		(IMAGE ClonEids 1323912-1325831, 1471368-1472903,			
1492104-1493255); NCI_CGAP_Lu5 pool 1 L1AM 3575-3582,		3851-3854 (IMAGE ClonEids 1414920-1417991, 1502904-1522439;			
); NCI_CGAP_GC4 pool 1 L1AM 3164-3167, 3716-3720,		3733-3735 (IMAGE ClonEids 1257096-1258631, 1469064-1470983,			
1475592-1476743); NCI_CGAP_Pr22 pool 1 L1AM 2457-2459,		2758-2759, 3082-3068 (IMAGE ClonEids 985608-986759			
1101192-1101959, 1217928-1220615); NCI_CGAP_Co10 pool 1		L1AM 2644-2653, 2871-2872 (IMAGE ClonEids 1057416-1061255			
, 1144584-1145351). (10% of the driver population), plus a		pool of 3,840 arrayed clones from NCI_CGAP_Sub1 (IMAGE			
ClonEids 2708616-2710535) and NCI_CGAP_Sub2 (IMAGE		ClonEids 2710536-2712455) (10% of the driver population			
); plus a pool of 11,136 clones from NCI_CGAP_Sub3 (IMAGE		ClonEids 2712456-2723591) (10% of the driver population),			
plus a pool of 5,472 clones from NCI_CGAP_Sub4 (IMAGE		ClonEids 2723592-2728969) (70% of the driver population).			
Subtraction was performed as previously described (Bonaldo		Lennon & Soares (1996): Normalization and Subtraction:			
Two Approaches To Facilitate Gene Discovery. Genome		Research 6, 791-806.			
TAG_Lib=NCI_CGAP_Lym2		TAG_TISSUE=Lymph node			
TAG_SEQ=AAATG"					
BASE COUNT		114 a	111 c	84 g	155 t
ORIGIN					
Query Match		12.3%; Score 346; DB 10; Length 464;			
Best Local Similarity		100.0%; Pred. No. 0;			
Matches		346; Conservative	0; Mismatches	0; Indels	0; Gaps
QY	1361	CTTAAATGGGAGGAGGACTGTGAACAAGACACACCTTAATATGGTTGATGCTCGAAGT 1420			
Db	454	CTTAAATGGGAGGAGGACTGTGAACAAGACACCTTAATATGGTTGATGCTCGAAGT 395			
QY	1421	AGCAATCTTCTGGAAACGCAAACTCTTTTAAAGAAAGTCCCTTAATTTAGAACACCCACA 1480			
Db	394	AGCAATCTTCTGGAAACGCAAACTCTTTTAAAGAAAGTCCCTTAATTTAGAACACCCACA 335			
QY	1481	AACTTCACATATCATTAATTAGCAAAATTTGCAAGAAAGTTGCTTTGAATGTTGGGAGAG 1540			
Db	334	AACTTCACATATCATTAATTAGCAAAATTTGCAAGAAAGTTGCTTTGAATGTTGGGAGAG 275			
QY	1541	GAAATCTATTGGCTCTCGTGGGTCTCTTTCATCTCAGAAATGCCAATCAGGTCAAGGTTT 1600			
Db	274	GAAATCTATTGGCTCTCGTGGGTCTCTTTCATCTCAGAAATGCCAATCAGGTCAAGGTTT 215			
QY	1601	GCTACATTTTGTATGTCTGTGATGCTTCTCCCAAGGATATTAATACTATATAGAGACTT 1660			
Db	214	GCTACATTTTGTATGTCTGTGATGCTTCTCCCAAGGATATTAATACTATATAGAGACTT 155			
QY	1661	GTGACAAAACAGAAATCATAAAGCTGCGAACCCGTGGCACACGCTCAT 1706			
Db	154	GTGACAAAACAGAAATCATAAAGCTGCGAACCCGTGGCACACGCTCAT 109			
RESULT 23					
A1468242/c					
LOCUS		A1468242	321 bp	mRNA	linear
DEFINITION		tg35g05.x1 Soares_NFL.T.GBC.S1 Homo sapiens cDNA clone			
		IMAGE:2110808 3' similar to contains Alu repetitive element., mRNA			
sequence.					
ACCESSION		A1468242			
VERSION		A1468242.1 GI:4330332			

polylinker; Site\_1: Not 1; Site\_2: Eco RI; NCI\_CGAP\_Sub5 is a subtracted library derived from NCI\_CGAP\_Sub4. The NCI\_CGAP\_Sub5 library had 3 million recombinants. A single-stranded DNA preparation of NCI\_CGAP\_Sub4 was used as a tracer in a subtractive hybridization with a driver comprising: the IMAGE pool (NCI\_CGAP\_Kid3 pool 1 IJLJAM 3334-3337, 3682-3683, 3798-3803 (IMAGE Clones 132376-132391, 1456008-1456775, 1500552-1502855); NCI\_CGAP\_Kid5 pool 1 LLAM 3338-3342, 3722-3725, 3776-3778 (IMAGE Clones 1323912-1325831, 1471368-1472503, 1492104-1493255); NCI\_CGAP\_Lu5 pool 1 LLAM 3575-3582, 3581-3584 (IMAGE Clones 1414920-1417991, 1520904-1522439); NCI\_CGAP\_GC4 pool 1 LLAM 3164-3167, 3716-3720, 3733-3735 (IMAGE Clones 1257096-1258631, 1469064-1470983, 1475592-1476743); NCI\_CGAP\_P+22 pool 1 LLAM 2457-2459, 2758-2759, 3062-3068 (IMAGE Clones 985608-986759, 1101192-1101959, 1217928-1220615); NCI\_CGAP\_Co10 pool 1 LLAM 2644-2653, 2871-2872 (IMAGE Clones 1037416-1061255, 1144584-1145351). (10% of the driver population), plus a pool of 3,840 arrayed clones from NCI\_CGAP\_Sub1 (IMAGE Clones 2708616-2710535) and NCI\_CGAP\_Sub2 (IMAGE Clones 2710536-2712455) (10% of the driver population), plus a pool of 11,136 clones from NCI\_CGAP\_Sub3 (IMAGE Clones 2712456-2723591) (10% of the driver population), plus a pool of 5,472 clones from NCI\_CGAP\_Sub4 (IMAGE Clones 2723592-2728969) (70% of the driver population). Subtraction was performed as previously described [Bonaldi, Lennon & Soares (1996): Normalization and Subtraction: Two Approaches To Facilitate Gene Discovery. Genome Research 6, 791-806.].

TAG\_LIB=NCI\_CGAP\_Lym2  
TAG\_TISSUE=Lymph node  
TAG\_SEQ=AAATG"

BASE COUNT 114 a 111 c 84 g 155 t  
ORIGIN

Query Match 12.3%; Score 346; DB 10; Length 464;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1361 CTTAAATGGCAGGAGGACTGTGAACAACACACCTTAATTAATGGTGTGATCTGAAGT 1420  
Db 454 CTTAAATGGCAGGAGGACTGTGAACAACACACCTTAATTAATGGTGTGATCTGAAGT 395

QY 1421 AGCAATCTTCTGGAAACGCAAACTCTTTTAAGGAAGTCCCTTAATTTAGAAACACCCACA 1480  
Db 394 AGCAATCTTCTGGAAACGCAAACTCTTTTAAGGAAGTCCCTTAATTTAGAAACACCCACA 335

QY 1481 AACTTCACATATCATATATTAGCAACAATTTGGAAGGAAGTTGCTTGAATGTTGGGAGAG 1540  
Db 334 AACTTCACATATCATATATTAGCAACAATTTGGAAGGAAGTTGCTTGAATGTTGGGAGAG 275

QY 1541 GAAATCTATTGGCTCTCGTGGGTCTCTTCATCTCAGAAATGCCAATCAGGTCAAGTTT 1600  
Db 274 GAAATCTATTGGCTCTCGTGGGTCTCTTCATCTCAGAAATGCCAATCAGGTCAAGTTT 215

QY 1601 GCTACATTTGTATGTGTGTGATGCTTCTCCCAAAGGTATATTAACTATATAGAGAGTT 1660  
Db 214 GCTACATTTGTATGTGTGTGATGCTTCTCCCAAAGGTATATTAACTATATAGAGAGTT 155

QY 1661 GTCAAAACAGATGATAAAGCTGCCAACCGTGGCACACGCTCAT 1706  
Db 154 GTCAAAACAGATGATAAAGCTGCCAACCGTGGCACACGCTCAT 109

RESULT 23  
AI468242/c  
LOCUS  
DEFINITION  
IMAGE:2110808 3' similar to contains Alu repetitive element.; mRNA  
sequence.  
ACCESSION  
AI468242  
VERSION  
AI468242.1 GI:4330332



KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1. (bases 1 to 321)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 1095 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 304.  
FEATURES  
source Location/Qualifiers  
1. .321  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2110808"  
/clone\_lib="Soares\_NFL\_T\_GBC\_S1"  
/lab\_host="DH10B"  
/notes="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with  
a modified polylinker; Site\_1: Not 1; Site\_2: Eco RI;  
Equal amounts of plasmid DNA from three normalized  
libraries (fetal lung NbHL19W, testis NHT, and B-cell  
NCI-CGAP\_GCB1) were mixed, and ss circles were made in  
vitro. Following HAP purification, this DNA was used as  
tracer in a subtractive hybridization reaction. The driver  
was PCR-amplified cDNAs from pools of 5,000 clones made  
from the same 3 libraries. The pools consisted of  
I.M.A.G.E. clones 297480-302087, 682632-687239,  
726408-728711, and 729096-731399. Subtraction by Bento  
Soares and M. Fatima Bonaldo."  
BASE COUNT 84 a 81 c 56 g 100 t  
ORIGIN  
Query Match 11.4%; Score 321; DB 9; Length 321;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 321; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1488 CATATCATATAATAGCAACAATTTGGAAGGAGTTGCTTGAATCTTGGGAGAGGAAATC 1547  
Db 321 CATATCATATAATAGCAACAATTTGGAAGGAGTTGCTTGAATCTTGGGAGAGGAAATC 262  
Qy 1548 TATTGGCTCTCGTGGGTCTTTCATCTCAGAAATGCCAATCAGGTCAAGTTTGTACAT 1607  
Db 261 TATTGGCTCTCGTGGGTCTTTCATCTCAGAAATGCCAATCAGGTCAAGTTTGTACAT 202  
Qy 1608 TTTGTATGTGTGTCTCTCCAAAGGTATATTAACTATATAAGAGAGTTGTGACAA 1667  
Db 201 TTTGTATGTGTGTCTCTCCAAAGGTATATTAACTATATAAGAGAGTTGTGACAA 142  
Qy 1668 AACAGAATGATAAGCTGGCAACCGTGGCACACGCTCATAGTTCTAGCTTCITGGGAGGT 1727  
Db 141 AACAGAATGATAAGCTGGCAACCGTGGCACACGCTCATAGTTCTAGCTTCITGGGAGGT 82  
Qy 1728 TGAGGAGGAGGAGTGGCTTGAACACAGAGTTTCAAGGCCAGCCTGGGCAACATAACAAGA 1787  
Db 81 TGAGGAGGAGGAGTGGCTTGAACACAGAGTTTCAAGGCCAGCCTGGGCAACATAACAAGA 22  
Qy 1788 TCCTGTCTCTCAAAAAA 1808  
Db 21 TCCTGTCTCTCAAAAAA 1

RESULT 24  
AW768364  
LOCUS hK73004.x1 NCI\_CGAP\_Lym12 Homo sapiens cdna clone IMAGE:3002311 3',  
DEFINITION mRNA sequence.  
413 bp mRNA linear EST 04-MAY-2000

AW768364  
VERSION AW768364.1 GI:7700383  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1. (bases 1 to 413)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Life Technologies catalog #: 11547-015  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL, send email to:  
info@image.llnl.gov  
Seq primer: -40UP from Gibco  
High quality sequence stop: 411.  
FEATURES  
source Location/Qualifiers  
1. .413  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3002311"  
/clone\_lib="NCI\_CGAP\_Lym12"  
/tissue\_type="lymphoma, follicular mixed small and large  
cell"  
/lab\_host="DH10B"  
/notes="Organ: lymph node; Vector: pCMV-SPORT6; Site\_1:  
SalI; Site\_2: NotI; Cloned unidirectionally. Primer:  
Oligo dT. Average insert size 1.25 kb. Life Technologies  
catalog #: 11547-015"  
BASE COUNT 127 a 72 c 94 g 118 t 2 others  
ORIGIN  
Query Match 11.1%; Score 312; DB 10; Length 413;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 312; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1282 GTACAAAGAAGATTGTTATGGTGGGATGGAGGTATAGACCATGATGGTCACTTCA 1341  
Db 43 GTACAAAGAAGATTGTTATGGTGGGATGGAGGTATAGACCATGATGGTCACTTCA 102  
Qy 1342 AGCTACTTTAATAAAGATCTTAAATGGCAGGAGACTGTGAACAAGACACCCCTAATA 1401  
Db 103 AGCTACTTTAATAAAGATCTTAAATGGCAGGAGACTGTGAACAAGACACCCCTAATA 162  
Qy 1402 ATGGGTTGATGTCGTAAGTAGCAAACTCTTCTGGAAACGCAAACTCTTTTAAAGGAAGTCCC 1461  
Db 163 ATGGGTTGATGTCGTAAGTAGCAAACTCTTCTGGAAACGCAAACTCTTTTAAAGGAAGTCCC 222  
Qy 1462 TAATTTAGAAACACCCACAACTTCACATATCATATAATTAGCAACAATTTGGAAGGAGTT 1521  
Db 223 TAATTTAGAAACACCCACAACTTCACATATCATATAATTAGCAACAATTTGGAAGGAGTT 282  
Qy 1522 GCTTTGAATGTTGGGAGAGAGAAATCTATTGGCTCTCGTGGGTCTCTTCATCTCAGAAAT 1581  
Db 283 GCTTTGAATGTTGGGAGAGAGAAATCTATTGGCTCTCGTGGGTCTCTTCATCTCAGAAAT 342  
Qy 1582 GCCAATCAGGTC 1593  
Db 343 GCCAATCAGGTC 354  
RESULT 25  
AI249134/c  
LOCUS AI249134  
DEFINITION qh79c07.x1 Soares\_fetal\_liver\_spleen\_INFLS\_S1 Homo sapiens cdna  
clone IMAGE:1850892 3' similar to contains Alu repetitive element  
;contains element L1 repetitive element ;, mRNA sequence.  
ACCESSION AI249134

```
VERSION      AI249134.1  GI:3844531
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 363)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
JOURNAL      Unpublished (1997)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             This clone is available royalty-free through LLNL; contact the
             IMAGE Consortium (info@image.llnl.gov) for further information.
             Insert length: 1487, Std Error: 0.00
             Seq primer: -40UP from Gibco
             High quality sequence stop: 360.
FEATURES     Location/Qualifiers
             source          1..363
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clones="IMAGE:1850892"
             /clone_lib="Soares_fetal_liver_spleen_LNPLS_S1"
             /sex="male"
             /dev_stage="20 week-post conception fetus"
             /lab_host="DH10B (ampicillin resistant)"
             /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
             with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
             This is a subtracted version of the original Soares fetal
             liver spleen LNPLS library. 1st strand cDNA was primed
             with a Pac I - oligo(dT) primer [5',
             AACTGGAAGATTAATTAAGATCTTTTGTGTTTGTGTTT 3'],
             double-stranded cDNA was ligated to Eco RI adaptors
             (Pharmacia), digested with Pac I and cloned into the Pac I
             and Eco RI sites of the modified pT7T3 vector. Library
             went through one round of normalization. Library
             constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT   95 a 85 c 68 g 115 t
ORIGIN
Query Match 10.9%; Score 308; DB 9; Length 363;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 308; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1500 AGCAACAATTTGGAAGGAGTTGCTTGAATGCTTGGGAGAGGAGAAATCTATTGGCTCTCG 1559
Db 308 AGCAACAATTTGGAAGGAGTTGCTTGAATGCTTGGGAGAGGAGAAATCTATTGGCTCTCG 249
Qy 1560 TGGGCTCTTCTATCTCAGAAATGCCAATCAGGTCAAGGTTTGCTTACATTTTGTATGTG 1619
Db 248 TGGGCTCTTCTATCTCAGAAATGCCAATCAGGTCAAGGTTTGCTTACATTTTGTATGTG 189
Qy 1620 TGATGCTTCTCCCAAGGATATTAATCTATATAGAGAGTTGTGACAAAACAGATGATA 1679
Db 188 TGATGCTTCTCCCAAGGATATTAATCTATATAGAGAGTTGTGACAAAACAGATGATA 129
Qy 1680 AAGCTGGCAACCGTGGCACACGCTCATAGTCTCTTGGGAGGTTGAGGAGGAGG 1739
Db 128 AAGCTGGCAACCGTGGCACACGCTCATAGTCTCTTGGGAGGTTGAGGAGGAGG 69
Qy 1740 ATGGCTTTGAACACAGGTTTCAAGGCCAGCGCTGGGCAACATPACAAAGATCCTGCTCTCA 1799
Db 68 ATGGCTTTGAACACAGGTTTCAAGGCCAGCGCTGGGCAACATPACAAAGATCCTGCTCTCA 9
Qy 1800 AAAAAA 1807
Db 8 AAAAAA 1

RESULT 26
BF664355 LOCUS          956 bp      mRNA      linear      EST 21-DEC-2000
```

---

```
DEFINITION  602146013F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4309707 5',
             mRNA sequence.
ACCESSION   BF664355
VERSION     BF664355.1  GI:11938351
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 956)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLCM183 row: i column: 04
             High quality sequence stop: 725.
FEATURES     Location/Qualifiers
             source          1..956
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clones="IMAGE:4309707"
             /clone_lib="NIH_MGC_48"
             /tissue_type="primary B-cells from tonsils (cell line)"
             /lab_host="DH10B (phage-resistant)"
             /note="Organ: B-cells; Vector: pOTB7; Site_1: XhoI;
             Site_2: EcoRI; cDNA made by oligo-dT priming.
             Directionally cloned into EcoRI/XhoI sites using the
             following 5' adaptor: GCCACAG(G). Size-selected >500bp
             for average insert size 1.8kb. Library constructed by Ling
             Hong in the laboratory of Gerald M. Rubin (University of
             California, Berkeley) using ZAP-cDNA synthesis kit
             (Stratagene) and Superscript II RT (Life Technologies).
             Note: This is a NIH_MGC Library."
BASE COUNT   277 a 183 c 219 g 277 t
ORIGIN
Query Match 10.5%; Score 297; DB 12; Length 956;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 417; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
Qy 794 ACAGAAAAATATTATATACGACTCTTTAAAAAGATCTATGCTTGAANAATAGAGAAG 853
Db 162 ACAGAAAAATATTATATACGACTCTTTAAAAAGATCTATGCTTGAANAATAGAGAAG 221
Qy 854 AACACAGTCTGGCCAGGAGCTGCTGCAATTGTGTCAGTTTTCGAATGCAACATTGTCCC 913
Db 222 AACACAGTCTGGCCAGGAGCTGCTGCAATTGTGTCAGTTTTCGAATGCAACATTGTCCC 281
Qy 914 CTACTGGGAATAACAGAACTGCAGGACCTGGGAGCATCCTAAAGTGTCAAGCTTTTCTA 973
Db 282 CTACTGGGAATAACAGAACTGCAGGACCTGGGAGCATCCTAAAGTGTCAAGCTTTTCTA 341
Qy 974 TGACTTTTAGTAGATGAGAGCAGAGGTAGATCCTTAAAAAGCATGCTGAGAGATCAA 1033
Db 342 TGACTTTTAGTAGATGAGAGCAGAGGTAGATCCTTAAAAAGCATGCTGAGAGATCAA 401
Qy 1034 ATGTTTTTATATCAACATCCTTTTATTATTATTTTCTATT -TCAGTTTAACAGTGGTGTAGT 1092
Db 402 ATGTTTTTATATCAACATCCTTTTATTATTATTATTATTATTAGTTTACAGTGGTGTAGT 461
Qy 1093 GATAGATTTTCTATTCTTTTCCCTTTGACGTTTACTTTTCAAGTAACACAACTCTTCCAT 1152
Db 462 GATAGATTTTCTATTCTTTTCCCTTTGACGTTTACTTTTCAAGTAACACAACTCTTCCAT 521
Qy 1153 CAGGCCATGATCTATAGGACCTCCTTAATGAGAGTATCTGGGGTATTGTGACCCCAAC 1210
```



Db 157 TACTTTCAAGTAACACAACTCTTCCATCAGGCCATGATCTATAGGACCTCTCTAATGAGA 98

QY 1185 GTATCTGGTGATTTGAGCCCAACCATCTCTCCAAAGCATTAA 1231

Db 97 GTATCTGGTGATTTGAGCCCAACCATCTCTCCAAAGCATTAA 51

RESULT 29

AA504653/c

LOCUS

DEFINITION

aa63f11.s1 NCI\_CGAP\_GCB1 Homo sapiens cDNA clone IMAGE:825645 3'

similar to contains Alu repetitive element; , mRNA sequence.

ACCESSION

AA504653

VERSION

AA504653.1 GI:2240813

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 332)

AUTHORS

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE

National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL

Unpublished (1997)

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue procurement: Louis M. Staudt, M.D., David Allman, Ph.D., Gerald Marti, M.D.

cDNA library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)

Seq primer: -40m13 fwd. Et from Amersham

High quality sequence stop: 294.

Location/Qualifiers

1. 332

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:825645"

/clone\_lib="NCI\_CGAP\_GCB1"

/tissue\_type="germinal center B cell"

/lab\_host="DH10B"

/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was prepared from human tonsillar cells enriched for germinal center B cells by flow sorting (CD20+, IgD+), provided by Dr. Louis M. Staudt (NCI), Dr. David Allman (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was primed with a Not I - oligo(dT) primer

15'-NGTACCAATCTGAAGTGGAGCGCGCCCTATTCTTTTCTTTT-3'

1. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT

86 a 82 c 61 g 103 t

ORIGIN

Query Match 10.0%; Score 281; DB 9; Length 332;

Best Local Similarity 99.7%; Pred. No. 0;

Matches 331; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1476 CCACAACTTCATATCATATATAGCAACAAATTTGGAAGGAGTTGTTGAATGTGGG 1535

Db 332 CCACAACTTCATATCATATATAGCAACAAATTTGGAAGGAGTTGTTGAATGTGGG 273

QY 1536 GAGAGAAAATCTATTGGCTCGTGGGTCTCTTCATCTCAGAAATGCCAATCAGGTCAA 1595

Db 272 GAGAGAAAATCTATTGGCTCGTGGGTCTCTTCATCTCAGAAATGCCAATCAGGTCAA 213

QY 1596 GGTTCGTACATTTTGTATGTGTGTGTCCTCCCAAGAGTATATTAACATATATAAGA 1655

Db 212 GGTTCGTACATTTTGTATGTGTGTGTCCTCCCAAGAGTATATTAACATATATAAGA 153

QY 1656 GAGTTGTGACAAACACAGATGATAAGCTGCGAACCGTGGCACACGCTCATAGTTCTAGC 1715

Db 152 GAGTTGTGACAAACACAGATGATAAGCTGCGAACCGTGGCACACGCTCATAGTTCTAGC 93

QY 1716 TCCTTGGGAGGTTGAGGAGGAGGATGCTTCAACACACAGGTTCTCAAGCCAGCCCTGGGC 1775

Db 92 TCCTTGGGAGGTTGAGGAGGAGGATGCTTCAACACACAGGTTCTCAAGCCAGCCCTGGGC 33

QY 1776 AACATAACAAGATCCTGCTCTCTCAAAAAAAA 1807

Db 32 AACATAACAAGATCCTGCTCTCTCAAAAAAAA 1

RESULT 30

BG398364

LOCUS

DEFINITION

602439832F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4566366 5', mRNA sequence.

ACCESSION

BG398364

VERSION

BG398364.1 GI:13291812

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 928)

AUTHORS

NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Louis M. Staudt, M.D., Ph.D.

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Plate: LLCM1284 row: o column: 07

High quality sequence stop: 650.

Location/Qualifiers

1. 928

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:4566366"

/clone\_lib="NIH\_MGC\_48"

/tissue\_type="primary B-cells from tonsils (cell line)"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: B-cells; Vector: pOTB7; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT

280 a 163 c 222 g 263 t

ORIGIN

Query Match 9.9%; Score 279; DB 12; Length 928;

Best Local Similarity 99.6%; Pred. No. 0;

Matches 519; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY 665 TTGGGACTTTGATAGCAACTTCCAGGAATGTCACACAGCATCAATATCTCTGCTGAAGA 724

Db 97 TTGGGACTTTGATAGCAACTTCCAGGAATGTCACACAGCATCAATATCTCTGCTGAAGA 156

QY 725 CAGTGGATAAAAAACAGTCCTTCAAGTCTCTCTGTTTATTCTTCAACATCTCACTTTC 784

```

|||||
157 CAGTGGATAAAAACAGTCTCTCAAGTCTCTCTGTTTATTTCTTCAACTCTCACTTTC 216
|||||
785 TTAGAGTTTACAGAAAAATATTTATATACGACTCTTTTAAAGAGATCTATGCTTCAAAA 844
|||||
217 TTAGAGTTTACAGAAAAATATTTATATACGACTCTTTTAAAGAGATCTATGCTTCAAAA 276
|||||
845 TAGAGAAGAACACAGGCTGGCCAGGAGCGTCTGCAATTTGGTGCAGTTTTGAATGCAA 904
|||||
277 TAGAGAAGAACACAGGCTGGCCCA-GGAGCTGCTGCAATTTGGTGCAGTTTTGAATGCAA 335
|||||
905 CATTGTCCTCTACTGGATACAGAACTGCAGGACCTGGAGCATCTTAAAGTGTCAAC 964
|||||
336 CATTGTCCTCTACTGGATACAGAACTGCAGGACCTGGAGCATCTTAAAGTGTCAAC 395
|||||
965 GTTTTTCTATGACTTTTAGTAGGATGAGACGAGAGGTAGATCCTTAAAGCATGGTGA 1024
|||||
396 GTTTTTCTATGACTTTTAGTAGGATGAGACGAGAGGTAGATCCTTAAAGCATGGTGA 455
|||||
1025 GAGGATCAAAATG-TTTTTATATCAACATCCTTTATTATTGATTCATTTGAGTTAAACAGT 1083
|||||
456 GAGGATCAAAATGTTTTTATATCAACATCCTTTATTATTGATTCATTTGAGTTAAACAGT 515
|||||
1084 GGTGTTAGTAGATAGATTTTCTATTCTTTTCCCTTGACGTTTACTTTTCAAGTAACACAA 1143
|||||
516 GGTGTTAGTAGATAGATTTTCTATTCTTTTCCCTTGACGTTTACTTTTCAAGTAACACAA 575
|||||
1144 CTCTTCCATCAGGCCATGATCTATAGGACCTCTCTAATGAGA 1184
|||||
576 CTCTTCCATCAGGCCATGATCTATAGGACCTCTCTAATGAGA 616

```

```

RESULT 31
AW134750
LOCUS
DEFINITION
UI-H-B11-abq-d-11-0-UI.sl NCI-CGAP_Sub3 Homo sapiens cDNA clone
IMAGE:2712644 3', mRNA sequence.
AW134750
NCI34750.1 GI:6138296
EST.
SOURCE
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 454)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov

```

```

The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. cDNA library Preparation: M.B. Soares Lab Clone distribution:
NCI-CGAP clone distribution information can be found through the
I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html The following repetitive
elements were found in this cDNA sequence: 1-21,
>At_rich#Low_complexity
Seq primer: M13 Forward
POLYA=Yes.
Location/Qualifiers
1..454
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2712644"
/clone_lib="NCI-CGAP_Sub3"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The
NCI-CGAP_Sub3 library is a subtracted library derived from
the NCI-CGAP_Sub1 library, which is a subtracted library

```

```

FEATURES
source
1..454
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2712644"
/clone_lib="NCI-CGAP_Sub3"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The
NCI-CGAP_Sub3 library is a subtracted library derived from
the NCI-CGAP_Sub1 library, which is a subtracted library

```

```

derived from BI. BI constitutes a mixture of 21
normalized or subtracted NCI-CGAP libraries: NCI-CGAP_Co4
, NCI-CGAP_Pr22, NCI-CGAP_Pr28, NCI-CGAP_Co10,
NCI-CGAP_Co16, NCI-CGAP_Kid5, NCI-CGAP_Kid12,
NCI-CGAP_Kid3, NCI-CGAP_Kid11, NCI-CGAP_Lym2,
NCI-CGAP_Brn2, NCI-CGAP_Co8, NCI-CGAP_CLL1, NCI-CGAP_Lei2,
NCI-CGAP_Brn23, NCI-CGAP_Lu5, NCI-CGAP_Lu24,
NCI-CGAP_Brn25. These 21 libraries were pooled and a
single-stranded DNA preparation of the resulting mixture
was used as a tracer in a subtractive hybridization with
a driver whose composition is detailed below:
NCI-CGAP_Kid3 pool 1 LLAM 3334-3337, 3682-3683,
3798-3803 (IMAGE CloneIDs 1322376-1323911, 1456008-1456775
, 1500552-1502855); NCI-CGAP_Kid5 pool 1 LLAM 3338-3342
, 3722-3725, 3776-3778 (IMAGE CloneIDs 1323912-1325831,
1471368-1472903, 1492104-1493255); NCI-CGAP_Lu5 pool 1,
LLAM 3575-3582, 3851-3854 (IMAGE CloneIDs 1414920-1417991,
1520904-1522439); NCI-CGAP_GC4 pool 1 LLAM 3164-3167,
3716-3720, 3733-3735 (IMAGE CloneIDs 1257096-1258631,
1469064-1470983, 1475592-1476743); NCI-CGAP_Pr22 pool 1
LLAM 2457-2459, 2758-2759, 3062-3068 (IMAGE CloneIDs
985608-986759, 1101192-1101959, 1217928-1220615);
NCI-CGAP_Co10 pool 1 LLAM 2644-2653, 2871-2872 (IMAGE
CloneIDs 1057416-1061255, 1144584-1145351). Subtraction
was performed as previously described (Bonaldi, Lennon &
Soares (1996): Normalization and Subtraction: Two
Approaches To Facilitate Gene Discovery. Genome Research
6, 791-806.
TAG_LIB=NCI-CGAP_GC4
TAG_TISSUE=germ cell
TAG_SEQ=AAATC"
BASE COUNT 136 a 105 c 68 g 145 t
ORIGIN
Query Match 9.9%; Score 278; DB 10; Length 454;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 278; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 622 GCCCTGTATGAGTTGATGACTACGACGACGATTTTCGCTCTTGGGACTTTGATAGCA 681
|||||
Db 177 GCCCTGTATGAGTTGATGACTACGACGACGATTTTCGCTCTTGGGACTTTGATAGCA 236
|||||
QY 682 ACTTCCAGGAATGTCACACACGATGAATATCTCTGCTGAAGACAGTGGATAAAACAG 741
|||||
Db 237 ACTTCCAGGAATGTCACACACGATGAATATCTCTGCTGAAGACAGTGGATAAAACAG 296
|||||
QY 742 TCCTTCAAGTCTTCTCTGTTTTTTTATTTCTTCAACTCTCACTTCTTTAGAGTTTACAGAAA 801
|||||
Db 297 TCCTTCAAGTCTTCTCTGTTTTTTTATTTCTTCAACTCTCACTTCTTTAGAGTTTACAGAAA 356
|||||
QY 802 AATATTATATACGACTCTTAAAGAGATCTATGCTTCAAAATAGAGAGGACACAGG 861
|||||
Db 357 AATATTATATACGACTCTTAAAGAGATCTATGCTTCAAAATAGAGAGGACACAGG 416
|||||
QY 862 TCTGGCCAGGAGCGTCTGCAATTTGGTGCAGTTTTTCAA 899
|||||
Db 417 TCTGGCCAGGAGCGTCTGCAATTTGGTGCAGTTTTTCAA 454
|||||

```

```

RESULT 32
BG341819
LOCUS
DEFINITION
602463552F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4576188 5',
mRNA sequence.
ACCESSION
BG341819
VERSION
BG341819.1 GI:13148257
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 939)

```

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Plate: LLCMI288 row: h column: 13  
High quality sequence stop: 547.  
Location/Qualifiers  
1. 939  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4576188"  
/clone\_lib="NIH\_MGC\_48"  
/tissue.type="primary B-cells from tonsils (cell line)"  
/lab\_host="PH10B (phage-resistant)"  
/note="Organ: B-cells; Vector: pOTB7; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."  
Note: this is a NIH\_MGC library."  
BASE COUNT 253 a 196 c 222 g 268 t  
ORIGIN  
Query Match 7.3%; Score 207; DB 12; Length 939;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 207; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 665 TTGGGACTTTGATAGCAACTCCAGGAATGTCACACACGATGAATATCTCTGCTGAAGA 724  
Db 98 TTGGGACTTTGATAGCAACTCCAGGAATGTCACACACGATGAATATCTCTGCTGAAGA 157  
QY 725 CAGTGGATAAAACACAGTCTCTCAAGTCTCTCTGTTTATATCTCAACTCTCACTTTC 784  
Db 158 CAGTGGATAAAACACAGTCTCTCAAGTCTCTCTGTTTATATCTCAACTCTCACTTTC 217  
QY 785 TTAGAGTTTACAGAAAAATATTTATATACGACTCTTTAAAAAGATCTATGCTTCAAAA 844  
Db 218 TTAGAGTTTACAGAAAAATATTTATATACGACTCTTTAAAAAGATCTATGCTTGAAGA 277  
QY 845 TAGAGAAGGAACACAGGCTGGCCAGG 871  
Db 278 TAGAGAAGGAACACAGGCTGGCCAGG 304  
RESULT 33  
LOCUS BG170824  
DEFINITION 602323763F1 NIH\_MGC\_89 Homo sapiens cDNA clone IMAGE:4426959 5', mRNA sequence.  
ACCESSION BG170824  
VERSION BG170824.1 GI:12677527  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 517)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)

NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Plate: LLCMI288 row: h column: 13  
High quality sequence stop: 547.  
Location/Qualifiers  
1. 939  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4576188"  
/clone\_lib="NIH\_MGC\_48"  
/tissue.type="primary B-cells from tonsils (cell line)"  
/lab\_host="PH10B (phage-resistant)"  
/note="Organ: B-cells; Vector: pOTB7; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."  
Note: this is a NIH\_MGC library."  
BASE COUNT 253 a 196 c 222 g 268 t  
ORIGIN  
Query Match 7.3%; Score 207; DB 12; Length 939;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 207; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 665 TTGGGACTTTGATAGCAACTCCAGGAATGTCACACACGATGAATATCTCTGCTGAAGA 724  
Db 98 TTGGGACTTTGATAGCAACTCCAGGAATGTCACACACGATGAATATCTCTGCTGAAGA 157  
QY 725 CAGTGGATAAAACACAGTCTCTCAAGTCTCTCTGTTTATATCTCAACTCTCACTTTC 784  
Db 158 CAGTGGATAAAACACAGTCTCTCAAGTCTCTCTGTTTATATCTCAACTCTCACTTTC 217  
QY 785 TTAGAGTTTACAGAAAAATATTTATATACGACTCTTTAAAAAGATCTATGCTTCAAAA 844  
Db 218 TTAGAGTTTACAGAAAAATATTTATATACGACTCTTTAAAAAGATCTATGCTTGAAGA 277  
QY 845 TAGAGAAGGAACACAGGCTGGCCAGG 871  
Db 278 TAGAGAAGGAACACAGGCTGGCCAGG 304  
RESULT 33  
LOCUS BG170824  
DEFINITION 602323763F1 NIH\_MGC\_89 Homo sapiens cDNA clone IMAGE:4426959 5', mRNA sequence.  
ACCESSION BG170824  
VERSION BG170824.1 GI:12677527  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 517)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)

NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: ATCC  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Plate: LLCMI174 row: n column: 16  
High quality sequence stop: 517.  
Location/Qualifiers  
1. 517  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4426959"  
/clone\_lib="NIH\_MGC\_89"  
/tissue.type="hypermethylation, cell line"  
/lab\_host="PH10B (phage-resistant)"  
/note="Organ: kidney; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.3 kb. Library enriched for full-length clones and constructed by Life Technologies." Note: this is a NIH\_MGC Library."  
BASE COUNT 181 a 94 c 118 g 124 t  
ORIGIN  
Query Match 7.3%; Score 205; DB 12; Length 517;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 205; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1599 TTGCTACATTTTGTATGTGTGTGTCCTCTCCCAAGGTATATTAACATATATAAGAGAG 1658  
Db 283 TTGCTACATTTTGTATGTGTGTGTCCTCTCCCAAGGTATATTAACATATATAAGAGAG 342  
QY 1659 TTGTGACAAAACAGATGATAAAGCTGCGAACGCTGGCACAGCTCATAGTTCTAGCTGC 1718  
Db 343 TTGTGACAAAACAGATGATAAAGCTGCGAACGCTGGCACAGCTCATAGTTCTAGCTGC 402  
QY 1719 TTGGGAGGTTGAGGAGGAGGAGGCTTGAACACAGGTTTCAAGCCACGCTGGGCAAC 1778  
Db 403 TTGGGAGGTTGAGGAGGAGGAGGCTTGAACACAGGTTTCAAGCCACGCTGGGCAAC 462  
QY 1779 ATAACAAGATCTGCTCTCTCAAAAA 1803  
Db 463 ATAACAAGATCTGCTCTCTCAAAAA 487  
RESULT 34  
LOCUS AW444889  
DEFINITION UI-H-B13-ajz-c-08-0-UI.sl NCL\_CGAP\_Sub5 Homo sapiens cDNA clone IMAGE:2733327 3', mRNA sequence.  
ACCESSION AW444889  
VERSION AW444889.1 GI:6986651  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 215)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)  
Seq primer: M13 Forward

```
FEATURES
  source
    POLYA=Yes.
    Location/Qualifiers
      1..215
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="IMAGE:2733327"
        /clone_lib="NCI_CGAP_Sub5"
        /lab_host="DH10B (Life Technologies)"
        /note="vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; Site.1: Not I; Site.2: Eco RI; NCI_CGAP_Sub5
is a subtracted library derived from NCI_CGAP_Sub4. The
NCI_CGAP_Sub5 library had 3 million recombinants. A
single-stranded DNA preparation of NCI_CGAP_Sub4 was used
as a tracer in a subtractive hybridization with a driver
comprising: the IMAGE pool (NCI_CGAP_Kid3 pool 1 LLAM
3334-3337, 3682-3683, 3798-3803 (IMAGE Clones)
1322376-1323911, 1456008-1456775, 1500552-1502855);
NCI_CGAP_Kid5 pool 1 LLAM 3338-3342, 3722-3725, 3776-3778
(IMAGE Clones) 1323912-1325831, 1471368-1472903,
1492104-1493255); NCI_CGAP_Lu5 pool 1 LLAM 3575-3582,
3851-3854 (IMAGE Clones) 1414920-1417991, 1520904-1522439
); NCI_CGAP_GC4 pool 1 LLAM 3164-3167, 3716-3720,
3733-3735 (IMAGE Clones) 1257096-1258631, 1469064-1470983,
1475592-1476743); NCI_CGAP_Pr22 pool 1 LLAM 2457-2459,
2758-2759, 3062-3068 (IMAGE Clones) 983608-986759
, 1101192-1101959, 1217928-1220615); NCI_CGAP_Co10 pool 1
LLAM 2644-2653, 2871-2872 (IMAGE Clones) 1057416-1061255
, 1144584-1145351). (10% of the driver population), plus a
pool of 3,840 arrayed clones from NCI_CGAP_Sub1 (IMAGE
Clones) 2708616-2710533 and NCI_CGAP_Sub2 (IMAGE
Clones) 2710536-2712455) (10% of the driver population
), plus a pool of 11,136 clones from NCI_CGAP_Sub3 (IMAGE
Clones) 2712456-2723591) (10% of the driver population),
plus a pool of 5,472 clones from NCI_CGAP_Sub4 (IMAGE
Clones) 2723592-2728969) (70% of the driver population).
Subtraction was performed as previously described [Bonaldo
, Lennon & Soares (1996): Normalization and Subtraction:
Two Approaches To Facilitate Gene Discovery. Genome
Research 6, 791-806.
TAG_LIB=NCI_CGAP_Lym2
TAG_TISSUE=lymph node
TAG_SEQ=AAATG
BASE COUNT      67 a      50 g      61 t
ORIGIN
  Query Match
  Best Local Similarity 7.1%; Score 199; DB 10; Length 215;
  Matches 199; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1279 TTTGTACAAAAGAGATTGTTATGGTGGGATGAGGTATAGACCATGTCACCT 1338
|
|
|
Db 8 TTTGTACAAAAGAGATTGTTATGGTGGGATGAGGTATAGACCATGTCACCT 67
|
|
|
QY 1339 TCAGCTACTTATAAAGGATCTTAAATGGGAGGAGGACTGTGAACAAGACACCTA 1398
|
|
|
Db 68 TCAAGCTACTTATAAAGGATCTTAAATGGGAGGAGGACTGTGAACAAGACACCTA 127
|
|
|
QY 1399 ATAATGGGTGATGCTCAAGTACCAATCTTCTGGAACCAACCTCTTTTAAGGAGT 1458
|
|
|
Db 128 ATAATGGGTGATGCTCAAGTACCAATCTTCTGGAACCAACCTCTTTTAAGGAGT 187
|
|
|
QY 1459 CCCTAATTAGAAACACCC 1477
|
|
|
Db 188 CCCTAATTAGAAACACCC 206
|
|
|
RESULT 35
AA521498
LOCUS aa73h06.s1 NCI_CGAP_GCBI Homo sapiens cDNA clone IMAGE:826619 3',
DEFINITION mRNA sequence.
ACCESSION AA521498
VERSION AA521498.1 GI:2262041

KEYWORDS
SOURCE Homo sapiens
ORGANISM human.
EST.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
  1 (bases 1 to 463)
  NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
  Tumor Gene Index
  Unpublished (1997)
  JOURNAL
  COMMENT
    Contact: Robert Strausberg, Ph.D.
    Email: cgaps-r@mail.nih.gov
    Tissue Procurement: Louis M. Staudt, M.D., David Allman,
    Ph.D., Gerald Marti, M.D.
    cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
    Bonaldo, Ph.D.
    cDNA Library Arrayed by: Greg Lennon, Ph.D.
    DNA Sequencing by: Washington University Genome Sequencing Center
    Clone Distribution: NCI-CGAP clone distribution information can be
    found through the I.M.A.G.E. Consortium/LLNL at:
    www-bio.llnl.gov/bbrp/image/image.html
    Seq primer: -40ml3 fwd. ET from Amersham
    High quality sequence stop: 126.
  FEATURES
    source
      1..463
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="IMAGE:826619"
        /clone_lib="NCI_CGAP_GCBI"
        /tissue_type="germinal center B cell"
        /lab_host="DH10B"
        /note="vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI). cDNA synthesis was
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTCAAGTGGGAGGCGCTCATTTTTTTTTTTT-3'
]. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified p7T3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      135 a      94 c      94 g      140 t
ORIGIN
  Query Match
  Best Local Similarity 6.4%; Score 180; DB 9; Length 463;
  Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 730 GATAAAAACAGTCTTCAAGTCTTCTGTTTTTATCTTCAACTCTCACTTCTTGA 789
|
|
|
Db 1 GATAAAAACAGTCTTCAAGTCTTCTGTTTTTATCTTCAACTCTCACTTCTTGA 60
|
|
|
QY 790 GTTTACAGAAAAATATTTATATACGACTCTTTAAAAAGATCTATGCTTTGAAAATAGAG 849
|
|
|
Db 61 GTTTACAGAAAAATATTTATATACGACTCTTTAAAAAGATCTATGCTTTGAAAATAGAG 120
|
|
|
QY 850 AAGGAACACAGGTCTGGCCAGGACGCTGCTCAATTTGGTGCAGTTTGAATGCAACATTG 909
|
|
|
Db 121 AAGGAACACAGGTCTGGCCAGGACGCTGCTCAATTTGGTGCAGTTTGAATGCAACATTG 180
|
|
|
RESULT 36
AL581406/c
LOCUS AL581406 LTI_FL011_BCI Homo sapiens cDNA clone CS0DG003VB14 3 prime
DEFINITION , mRNA sequence.
ACCESSION AL581406
VERSION AL581406.1 GI:12948375
KEYWORDS EST.
SOURCE human.
```

ORGANISM	Homo sapiens									
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.									
AUTHORS	1 (bases 1 to 507)									
TITLE	Li, W. B., Gruber, C., Jessee, J., and Polayes, D.									
JOURNAL	Full-length cDNA libraries and normalization									
COMMENT	Unpublished (2001)									
CONTACT	Contact: Genoscope									
Genoscope	Genoscope - Centre National de Sequencage									
BP 191 91006 EVRY cedex - France										
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.										
FEATURES	Location/Qualifiers									
SOURCE	1..507									
/organism="Homo sapiens"										
/db_xref="taxon:9606"										
/clone="CS0DG003YB14"										
/clone_lib="LTI_FL011_BC1"										
/sex="male"										
/tissue_type="B cells from Burkitt lymphoma"										
/lab_host="DH10B"										
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"										
BASE COUNT	134 a	123 c	84 g	158 t	8 others					
ORIGIN										
Query Match	5.4%; Score 153; DB 9; Length 507;									
Best Local Similarity	100.0%; Pred. No. 0;									
Matches 153; Conservative	0; Mismatches	0; Indels	0; Gaps	0;						
QY	1412	GTCTGAAGTAGCAAAATCTTCTGGAAGCGCAAACTCTTTTAAAGGAAGTCCCTAATTTAGAA	1471							
Db	366	GTCTGAAGTAGCAAAATCTTCTGGAAGCGCAAACTCTTTTAAAGGAAGTCCCTAATTTAGAA	307							
QY	1472	ACACCCACAACCTTCACATATCATATAATTAGCAACAACTTGGAGGAAGTTCCTTGAATGT	1531							
Db	306	ACACCCACAACCTTCACATATCATATAATTAGCAACAACTTGGAGGAAGTTCCTTGAATGT	247							
QY	1532	TGGGGAGAGAAAATCTATTGGCTCTCGTGGGT	1564							
Db	246	TGGGGAGAGAAAATCTATTGGCTCTCGTGGGT	214							
RESULT 37										
AA879422/c										
LOCUS	AA879422									
DEFINITION	OJ91c11.s1 Soares_NFL_T-GBC.S1 Homo sapiens cDNA clone									
ACCESSION	IMAGE:1505684 3', mRNA sequence.									
VERSION	AA879422									
KEYWORDS	AA879422.1 GI:2988533									
SOURCE	EST.									
ORGANISM	human.									
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;										
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.										
1 (bases 1 to 268)										
NCI-CCAP	http://www.ncbi.nlm.nih.gov/ncicgap.									
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),										
Tumor Gene Index										
Unpublished (1997)										
Contact: Robert Strausberg, Ph.D.										
Email: ccapbs@email.nih.gov										
This clone is available royalty-free through LLNL ; contact the										
IMAGE Consortium (info@image.llnl.gov) for further information.										
Insert Length: 685 Std Error: 0.00										
Seq primer: -40ml3 fwd. ET from Amersham										
High quality sequence stop: 252.										



```

FEATURES
  source
    High quality sequence stop: 635.
    Location/Qualifiers
      1. .1140
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="IMAGE:4309563"
        /clone_lib="NIH_MGC_48"
        /tissue_type="primary B-cells from tonsils (cell line)"
        /lab_host="DH10B (phage-resistant)"
        /note="Organ: B-cells; Vector: pOTB7; Site_1: XhoI;
        Site_2: EcoRI; cDNA made by oligo-dT priming.
        Directionally cloned into EcoRI/XhoI sites using the
        following 5' adaptor: GGCACGAG(G). Size-selected >500bp
        for average insert size 1.8kb. Library constructed by Ling
        Hong in the laboratory of Gerald M. Rubin (University of
        California, Berkeley) using ZAP-cDNA synthesis kit
        (Stratagene, Berkeley) and Superscript II RT (Life Technologies).
        Note: this is a NIH_MGC Library."
BASE COUNT      361 a      252 c      277 g      249 t
ORIGIN
  Query Match      3.6%; Score 101; DB 12; Length 1140;
  Best Local Similarity 100.0%; Pred. No. 0;
  Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      868 CAGGGACGTGTCGAATTGGTCAGTTTGAATGCAACATTTGCCCTACTGGGAATAAC 927
          |||||||
Db      309 CAGGGACGTGTCGAATTGGTCAGTTTGAATGCAACATTTGCCCTACTGGGAATAAC 368

QY      928 AGAAGTGCAGGACCTGGGAGCATCTTAAGTGTCAACGTTT 968
          |||||||
Db      369 AGAAGTGCAGGACCTGGGAGCATCTTAAGTGTCAACGTTT 409

RESULT 39
AW182255/c
LOCUS
DEFINITION
  xJ72r10.x1 Soares_NFL_T_GBC.S1 Homo sapiens cDNA clone
  IMAGE:2662795 3' similar to contains Alu repetitive element
  ; contains element TAR1 repetitive element ; mRNA sequence.
ACCESSION
  AW182255
VERSION
  AW182255.1 GI:6450715
KEYWORDS
  EST.
SOURCE
  human.
  ORGANISM
    Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE
    1 (bases 1 to 364)
    NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
    National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
    Tumor Gene Index
    Unpublished (1997)
    Contact: Robert Strausberg, Ph.D.
    Email: cgapbs-r@mail.nih.gov
  JOURNAL
    This clone is available royalty-free through LML : contact the
    IMAGE Consortium (info@image.llnl.gov) for further information.
    Seq primer: -40UP from Gibco
    High quality sequence stop: 354.
  FEATURES
    Location/Qualifiers
      1. .364
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="IMAGE:2662795"
        /clone_lib="Soares_NFL_T_GBC_S1"
        /lab_host="DH10B"
        /note="Organ: pooled; Vector: p7T3D-Pac (Pharmacia) with
        a modified polylinker; Site_1: Not I; Site_2: Eco RI;
        Equal amounts of plasmid DNA from three normalized
        libraries (fetal lung NbHL19w, testis NHT, and B-cell
        NCI-CGAP.GC81) were mixed, and ss circles were made in
        vitro. Following HAP purification, this DNA was used as
        tracer in a subtractive hybridization reaction. The driver
        was PCR-amplified cDNAs from pools of 5,000 clones made
        from the same 3 libraries. The pools consisted of
        I.M.A.G.E. clones 297480-302087, 682632-687239,
        726408-728711, and 729096-731399. Subtraction by Bento
        Soares and M. Fatima Bonaldo."
BASE COUNT      94 a      86 c      76 g      108 t
ORIGIN
  Query Match      3.0%; Score 85; DB 10; Length 364;
  Best Local Similarity 100.0%; Pred. No. 0;
  Matches 85; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1724 AGTTTGAGGAGGAGGATGGCTTGAACACAGCAGGTGTTCAAGGCCAGCCTGGGCAACATAAC 1783
          |||||||
Db      85 AGTTTGAGGAGGAGGATGGCTTGAACACAGCAGGTGTTCAAGGCCAGCCTGGGCAACATAAC 26

QY      1784 AAGATCCCTGCTCTCAAAAAA 1808
          |||||||
Db      25 AAGATCCCTGCTCTCAAAAAA 1

RESULT 40
AA074342/c
LOCUS
DEFINITION
  zml5903.s1 Stratagene pancreas (#937208) Homo sapiens cDNA clone
  IMAGE:525748 3' similar to contains Alu repetitive element; mRNA
  sequence.
ACCESSION
  AA074342
VERSION
  AA074342.1 GI:1614210
KEYWORDS
  EST.
SOURCE
  human.
  ORGANISM
    Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE
    1 (bases 1 to 154)
    Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,
    Chissoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W., Hawkins
    , M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore
    , B., Morris, M., Parsons, J., Prange, C., Riffkin, L., Rohlfing, T.,
    Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevas, E.,
    Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Warra, M.
    Generation and analysis of 280,000 human expressed sequence tags
    Genome Res. 6 (9), 807-828 (1996)
  TITLE
    Contact: Wilson RK
  JOURNAL
    Washington University School of Medicine
  MEDLINE
    4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
    Tel: 314 286 1800
    Fax: 314 286 1810
    Email: est@watson.wustl.edu
  COMMENT
    WARNING: There is evidence that suggests that the 384-well parent
    plate of this clone contains both human and mouse derived clones.
    Thus, the origin of this clone is uncertain. This caution should be
    kept in mind should you use this clone.

This clone is available royalty-free through LML : contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40M13 fwd. from Amersham
High quality sequence stop: 126.
  FEATURES
    Location/Qualifiers
      1. .154
        /organism="Homo sapiens"
        /db_xref="GDB:391717"
        /db_xref="taxon:9606"
        /clone="IMAGE:525748"
        /clone_lib="Stratagene pancreas (#937208)"
        /lab_host="SOLR cells (kanamycin resistant)"
        /note="Organ: pancreas; Vector: pBluescript SK-; Site_1:
        EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
        Oligo dT. Panceratic adenocarcinoma cell line. Average
        insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor
        sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor sequence: 5'
        CTCGAGATTTTTTTTTTTT 3'"
BASE COUNT      28 a      53 c      39 g      34 t

```

ORIGIN		RESULT 42	
AV719004/c		AV719004	
Query Match		586 bp	
Best Local Similarity		mRNA	
Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		linear	
EST.		EST 16-OCT-2000	
QY 1983 GCACCTGTATCCCGAGCTACTTGGGAGGCTGAGGAGGAGAGTGCCTTGAACCCAGGAGG 2042		AV719004 GLC Homo sapiens cDNA clone GLCEJB05 5', mRNA sequence.	
Db 129 GCACCTGTATCCCGAGCTACTTGGGAGGCTGAGGAGGAGAGTGCCTTGAACCCAGGAGG 70		AV719004	
QY 2043 TGGAGGTTGCAGT 2055		AV719004	
Db 69 TGGAGGTTGCAGT 57		EST.	
TITLE		GI:10816156	
LOCUS		human.	
DEFINITION		Homo sapiens	
ACCESSION		Homo sapiens	
VERSION		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
KEYWORDS		Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
SOURCE		1 (bases 1 to 586)	
ORGANISM		Qian, B., Wu, T., Huang, Q., Kang, B., Gao, X., Xu, Z., Xiao, H.,	
REFERENCE		, Xu, X., Li, N., Peng, Y., Liu, F., Ou, J., Song, H., Cheng, Z., Zeng, L.,	
AUTHORS		Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Yang	
JOURNAL		, Y., Gu, Y., Chen, Z. and Han, Z.	
COMMENT		Homo sapiens cDNA GLC clones	
TITLE		Unpublished (2000)	
LOCUS		Contact: Zeguang Han	
DEFINITION		Chinese National Human Genome Center at Shanghai	
ACCESSION		351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai	
VERSION		201203, P. R. China	
KEYWORDS		Tel: 86-21-50801919(ex.45)	
SOURCE		Fax: 86-21-50801922	
ORGANISM		Email: hanzg@chgc.sh.cn	
REFERENCE		This clone is available at CHGC in Shanghai.	
AUTHORS		Location/Qualifiers	
JOURNAL		1..586	
COMMENT		/organism="Homo sapiens"	
TITLE		/db_xref="taxon:9606"	
LOCUS		/clone="GLCEJB05"	
DEFINITION		/clone_lib="GLC"	
ACCESSION		/tissue_type="corresponding non cancerous liver tissue"	
VERSION		/dev_stage="Adult"	
KEYWORDS		/lab_host="SOLR"	
SOURCE		/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:	
ORGANISM		XhoI"	
REFERENCE		BASE COUNT 139 a 144 c 154 g 149 t	
AUTHORS		ORIGIN	
JOURNAL		Query Match 2.6%; Score 72; DB 10; Length 586;	
COMMENT		Best Local Similarity 100.0%; Pred. No. 0;	
TITLE		Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
LOCUS		QY 1982 GCACCTGTATCCCGAGCTACTTGGGAGGCTGAGGAGGAGAGTGCCTTGAACCCAGGAG 2041	
DEFINITION		Db 116 GCACCTGTATCCCGAGCTACTTGGGAGGCTGAGGAGGAGAGTGCCTTGAACCCAGGAG 57	
ACCESSION		QY 2042 GTGGAGGTTGCAG 2053	
VERSION		Db 56 GTGGAGGTTGCAG 45	
KEYWORDS		TITLE	
SOURCE		Query Match 2.6%; Score 73; DB 12; Length 878;	
ORGANISM		Best Local Similarity 100.0%; Pred. No. 0;	
REFERENCE		Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
AUTHORS		QY 1982 GCACCTGTATCCCGAGCTACTTGGGAGGCTGAGGAGGAGAGTGCCTTGAACCCAGGAG 2041	
JOURNAL		Db 516 GCACCTGTATCCCGAGCTACTTGGGAGGCTGAGGAGGAGAGTGCCTTGAACCCAGGAG 457	
COMMENT		QY 2042 GTGGAGGTTGCAG 2054	
TITLE		Db 456 GTGGAGGTTGCAG 444	
LOCUS		BASE COUNT 197 a 257 c 246 g 178 t	
DEFINITION		ORIGIN	
ACCESSION		Query Match 2.6%; Score 73; DB 12; Length 878;	
VERSION		Best Local Similarity 100.0%; Pred. No. 0;	
KEYWORDS		Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
SOURCE		QY 1982 GCACCTGTATCCCGAGCTACTTGGGAGGCTGAGGAGGAGAGTGCCTTGAACCCAGGAG 2041	
ORGANISM		Db 516 GCACCTGTATCCCGAGCTACTTGGGAGGCTGAGGAGGAGAGTGCCTTGAACCCAGGAG 457	
REFERENCE		QY 2042 GTGGAGGTTGCAG 2054	
AUTHORS		Db 456 GTGGAGGTTGCAG 444	
JOURNAL		TITLE	
COMMENT		Query Match 2.6%; Score 73; DB 12; Length 878;	
TITLE		Best Local Similarity 100.0%; Pred. No. 0;	
LOCUS		Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
DEFINITION		QY 1982 GCACCTGTATCCCGAGCTACTTGGGAGGCTGAGGAGGAGAGTGCCTTGAACCCAGGAG 2041	
ACCESSION		Db 516 GCACCTGTATCCCGAGCTACTTGGGAGGCTGAGGAGGAGAGTGCCTTGAACCCAGGAG 457	
VERSION		QY 2042 GTGGAGGTTGCAG 2054	
KEYWORDS		Db 456 GTGGAGGTTGCAG 444	
SOURCE		TITLE	
ORGANISM		Query Match 2.6%; Score 73; DB 12; Length 878;	
REFERENCE		Best Local Similarity 100.0%; Pred. No. 0;	
AUTHORS		Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
JOURNAL		QY 1982 GCACCTGTATCCCGAGCTACTTGGGAGGCTGAGGAGGAGAGTGCCTTGAACCCAGGAG 2041	
COMMENT		Db 516 GCACCTGTATCCCGAGCTACTTGGGAGGCTGAGGAGGAGAGTGCCTTGAACCCAGGAG 457	
TITLE		QY 2042 GTGGAGGTTGCAG 2054	
LOCUS		Db 456 GTGGAGGTTGCAG 444	
DEFINITION		TITLE	
ACCESSION		Query Match 2.6%; Score 73; DB 12; Length 878;	
VERSION		Best Local Similarity 100.0%; Pred. No. 0;	
KEYWORDS		Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
SOURCE		QY 1982 GCACCTGTATCCCGAGCTACTTGGGAGGCTGAGGAGGAGAGTGCCTTGAACCCAGGAG 2041	
ORGANISM		Db 516 GCACCTGTATCCCGAGCTACTTGGGAGGCTGAGGAGGAGAGTGCCTTGAACCCAGGAG 457	
REFERENCE		QY 2042 GTGGAGGTTGCAG 2054	
AUTHORS		Db 456 GTGGAGGTTGCAG 444	
JOURNAL		TITLE	
COMMENT		Query Match 2.6%; Score 73; DB 12; Length 878;	
TITLE		Best Local Similarity 100.0%; Pred. No. 0;	
LOCUS		Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
DEFINITION		QY 1982 GCACCTGTATCCCGAGCTACTTGGGAGGCTGAGGAGGAGAGTGCCTTGAACCCAGGAG 2041	
ACCESSION		Db 516 GCACCTGTATCCCGAGCTACTTGGGAGGCTGAGGAGGAGAGTGCCTTGAACCCAGGAG 457	
VERSION		QY 2042 GTGGAGGTTGCAG 2054	
KEYWORDS		Db 456 GTGGAGGTTGCAG 444	
SOURCE		TITLE	
ORGANISM		Query Match 2.6%; Score 73; DB 12; Length 878;	
REFERENCE		Best Local Similarity 100.0%; Pred. No. 0;	
AUTHORS		Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
JOURNAL		QY 1982 GCACCTGTATCCCGAGCTACTTGGGAGGCTGAGGAGGAGAGTGCCTTGAACCCAGGAG 2041	
COMMENT		Db 516 GCACCTGTATCCCGAGCTACTTGGGAGGCTGAGGAGGAGAGTGCCTTGAACCCAGGAG 457	
TITLE		QY 2042 GTGGAGGTTGCAG 2054	
LOCUS		Db 456 GTGGAGGTTGCAG 444	
DEFINITION		TITLE	
ACCESSION		Query Match 2.6%; Score 73; DB 12; Length 878;	
VERSION		Best Local Similarity 100.0%; Pred. No. 0;	
KEYWORDS		Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
SOURCE		QY 1982 GCACCTGTATCCCGAGCTACTTGGGAGGCTGAGGAGGAGAGTGCCTTGAACCCAGGAG 2041	
ORGANISM		Db 516 GCACCTGTATCCCGAGCTACTTGGGAGGCTGAGGAGGAGAGTGCCTTGAACCCAGGAG 457	
REFERENCE		QY 2042 GTGGAGGTTGCAG 2054	
AUTHORS		Db 456 GTGGAGGTTGCAG 444	
JOURNAL		TITLE	
COMMENT		Query Match 2.6%; Score 73; DB 12; Length 878;	
TITLE		Best Local Similarity 100.0%; Pred. No. 0;	
LOCUS		Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
DEFINITION		QY 1982 GCACCTGTATCCCGAGCTACTTGGGAGGCTGAGGAGGAGAGTGCCTTGAACCCAGGAG 2041	
ACCESSION		Db 516 GCACCTGTATCCCGAGCTACTTGGGAGGCTGAGGAGGAGAGTGCCTTGAACCCAGGAG 457	
VERSION		QY 2042 GTGGAGGTTGCAG 2054	
KEYWORDS		Db 456 GTGGAGGTTGCAG 444	
SOURCE		TITLE	
ORGANISM		Query Match 2.6%; Score 73; DB 12; Length 878;	
REFERENCE		Best Local Similarity 100.0%; Pred. No. 0;	
AUTHORS		Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
JOURNAL		QY 1982 GCACCTGTATCCCGAGCTACTTGGGAGGCTGAGGAGGAGAGTGCCTTGAACCCAGGAG 2041	
COMMENT		Db 516 GCACCTGTATCCCGAGCTACTTGGGAGGCTGAGGAGGAGAGTGCCTTGAACCCAGGAG 457	
TITLE		QY 2042 GTGGAGGTTGCAG 2054	
LOCUS		Db 456 GTGGAGGTTGCAG 444	
DEFINITION		TITLE	
ACCESSION		Query Match 2.6%; Score 73; DB 12; Length 878;	
VERSION		Best Local Similarity 100.0%; Pred. No. 0;	
KEYWORDS		Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
SOURCE		QY 1982 GCACCTGTATCCCGAGCTACTTGGGAGGCTGAGGAGGAGAGTGCCTTGAACCCAGGAG 2041	
ORGANISM		Db 516 GCACCTGTATCCCGAGCTACTTGGGAGGCTGAGGAGGAGAGTGCCTTGAACCCAGGAG 457	
REFERENCE		QY 2042 GTGGAGGTTGCAG 2054	
AUTHORS		Db 456 GTGGAGGTTGCAG 444	
JOURNAL		TITLE	
COMMENT		Query Match 2.6%; Score 73; DB 12; Length 878;	
TITLE		Best Local Similarity 100.0%; Pred. No. 0;	
LOCUS		Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
DEFINITION		QY 1982 GCACCTGTATCCCGAGCTACTTGGGAGGCTGAGGAGGAGAGTGCCTTGAACCCAGGAG 2041	
ACCESSION		Db 516 GCACCTGTATCCCGAGCTACTTGGGAGGCTGAGGAGGAGAGTGCCTTGAACCCAGGAG 457	
VERSION		QY 2042 GTGGAGGTTGCAG 2054	
KEYWORDS		Db 456 GTGGAGGTTGCAG 444	
SOURCE		TITLE	
ORGANISM		Query Match 2.6%; Score 73; DB 12; Length 878;	
REFERENCE		Best Local Similarity 100.0%; Pred. No. 0;	
AUTHORS		Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
JOURNAL		QY 1982 GCACCTGTATCCCGAGCTACTTGGGAGGCTGAGGAGGAGAGTGCCTTGAACCCAGGAG 2041	
COMMENT		Db 516 GCACCTGTATCCCGAGCTACTTGGGAGGCTGAGGAGGAGAGTGCCTTGAACCCAGGAG 457	
TITLE		QY 2042 GTGGAGGTTGCAG 2054	
LOCUS		Db 456 GTGGAGGTTGCAG 444	
DEFINITION		TITLE	
ACCESSION		Query Match 2.6%; Score 73; DB 12; Length 878;	
VERSION		Best Local Similarity 100.0%; Pred. No. 0;	
KEYWORDS		Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
SOURCE		QY 1982 GCACCTGTATCCCGAGCTACTTGGGAGGCTGAGGAGGAGAGTGCCTTGAACCCAGGAG 2041	
ORGANISM		Db 516 GCACCTGTATCCCGAGCTACTTGGGAGGCTGAGGAGGAGAGTGCCTTGAACCCAGGAG 457	
REFERENCE		QY 2042 GTGGAGGTTGCAG 2054	
AUTHORS		Db 456 GTGGAGGTTGCAG 444	
JOURNAL		TITLE	
COMMENT		Query Match 2.6%; Score 73; DB 12; Length 878;	
TITLE		Best Local Similarity 100.0%; Pred. No. 0;	
LOCUS		Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	

JOURNAL  
MEDLINE  
COMMENT

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil

Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=4t2-RC5-BN0158-190  
400-031-G01&t3=2000-04-19&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 11  
High quality sequence stop: 251.  
Location/Qualifiers

FEATURES  
source

1..252  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="BN0158"  
/dev\_stage="Adult"  
/note="Organ: breast\_normal; Vector: puc18; Site\_1: SmaI;  
Site\_2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."  
53 a 67 c 60 g 72 t

BASE COUNT  
ORIGIN

Query Match 2.5%; Score 70; DB 10; Length 252;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1986 CCTGTAATCCCAGCTACTTGGAGGCTGAGCAGGAGAGATCGCTTTGAACCCAGGAGGTGG 2045  
|||||  
Db 181 CCTGTAATCCCAGCTACTTGGAGGCTGAGCAGGAGAGATCGCTTTGAACCCAGGAGGTGG 122  
QY 2046 AGGTTGCAGT 2055  
|||||  
Db 121 AGGTTGCAGT 112

RESULT 44  
BE012103/c

LOCUS RC5-BN1058-020500-031-G01 BN1058 Homo sapiens cDNA, mRNA sequence.  
DEFINITION BE012103  
ACCESSION BE012103  
VERSION BE012103.1 GI:8272336  
KEYWORDS EST.  
SOURCE human.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 253)  
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,  
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.

TITLE

Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags

JOURNAL  
MEDLINE  
COMMENT

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil

Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=4t2-RC5-BN1058-020  
500-031-G01&t3=2000-05-02&t4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 252.  
Location/Qualifiers

FEATURES  
source

1..253  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="BN1058"  
/dev\_stage="Adult"  
/note="Organ: breast\_normal; Vector: puc18; Site\_1: SmaI;  
Site\_2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."  
53 a 67 c 61 g 72 t

BASE COUNT  
ORIGIN

Query Match 2.5%; Score 70; DB 10; Length 253;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1986 CCTGTAATCCCAGCTACTTGGAGGCTGAGCAGGAGAGATCGCTTTGAACCCAGGAGGTGG 2045  
|||||  
Db 181 CCTGTAATCCCAGCTACTTGGAGGCTGAGCAGGAGAGATCGCTTTGAACCCAGGAGGTGG 122  
QY 2046 AGGTTGCAGT 2055  
|||||  
Db 121 AGGTTGCAGT 112

RESULT 45  
AA548692/c

LOCUS nk03901.sl NCI-CGAP\_Prl1 Homo sapiens cDNA clone IMAGE:1000944  
DEFINITION similar to contains Alu repetitive element; , mRNA sequence.  
ACCESSION AA548692  
VERSION AA548692.1 GI:2318974  
KEYWORDS EST.  
SOURCE human.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 318)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov

REFERENCE

1 (bases 1 to 318)  
AUTHORS Chuquiqui, M.D., Michael R. Emmert-Buck, M.D., Ph.D., Rodrigo F.  
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D.,  
DNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution Information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html  
Insert Length: 590 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 296.  
Location/Qualifiers

FEATURES  
source

1..318  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1000944"

```

/clone_lib="NCI_CGAP_Prill"
/sex="male"
/tissue_type="normal prostatic epithelial cells"
/lab_host="DH10B"
/notes="Organ: prostate; Vector: pAMP10; mRNA made from
normal prostatic epithelial cells, cDNA made by oligo-dT
priming. Non-directionally cloned. Size-selected on
agarose gel, average insert size 600 bp. Library made by
D. Krizman, NIH."
BASE COUNT      65 a      86 c      70 g      97 t
ORIGIN
|||||
Query Match      2.5%; Score 70; DB 9; Length 318;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1986 CCTGTAAATCCAGCTACTTGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAGGAGGTGG 2045
|||||
Db 137 CTTGTAAATCCAGCTACTTGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAGGAGGTGG 78
|||||
QY 2046 AGGTTGCAGT 2055
|||||
Db 77 AGGTTGCAGT 68

RESULT 46
LOCUS AV759459 334 bp mRNA linear EST 19-OCT-2000
DEFINITION MDS Homo sapiens cDNA clone MDSARB02 5', mRNA sequence.
ACCESSION AV759459
VERSION AV759459.1 GI:10917307
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 334)
AUTHORS Gu, J., Zhao, M., Huang, Q., Xu, X., Li, Y., Peng, Y., Song, H., Xiao, H.,
Gu, Y., Li, N., Qian, B., Liu, F., Ou, J., Gao, X., Cheng, Z., Xu, Z., Zeng
L., Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G.,
Yang, Y., Gao, G., Zhang, Q., Chen, S., Han, Z. and Chen, Z.
TITLE Homo sapiens cDNA MDS clones
JOURNAL Unpublished (2000)
COMMENT Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203. P. R. China
Tel: 86-21-50801919(ex. 45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
FEATURES
Location/Qualifiers
1..334
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MDSARB02"
/clone_lib="MDS"
/tissue_type="Bone marrow"
/cell_type="CD34+ hematopoietic stem/progenitor cell"
/lab_host="BM25.8"
/notes="Vector: pTriplEx2; Site_1: sfIIA; Site_2: sfIIB"
BASE COUNT      67 a      103 c      93 g      71 t
ORIGIN
|||||
Query Match      2.5%; Score 70; DB 10; Length 334;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1986 CCTGTAAATCCAGCTACTTGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAGGAGGTGG 2045
|||||
Db 86 CTTGTAAATCCAGCTACTTGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAGGAGGTGG 145
|||||
QY 2046 AGGTTGCAGT 2055

```

```

|||||
Db 146 AGGTTGCAGT 155

RESULT 47
LOCUS AA211885 346 bp mRNA linear EST 13-AUG-1997
DEFINITION zr90c03.s1 NCI_CGAP GCB1 Homo sapiens cDNA clone IMAGE:682948 3',
similar to contains Alu repetitive element.; mRNA sequence.
ACCESSION AA211885
VERSION AA211885.1 GI:1810530
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 346)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -41ml3 fwd. Et from Amersham
High quality sequence stop: 260.
FEATURES
Location/Qualifiers
1..346
/organism="Homo sapiens"
/db_xref="GDB:5586268"
/db_xref="taxon:9606"
/clone="IMAGE:682948"
/clone_lib="NCI_CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/notes="Vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI). Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - Oligo(dT) primer
[5'-TGTTACCAATCTGAAGTGGAGCGCGCTCATTTTTTTTTTTT-3'.
]. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified p7T3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      76 a      90 c      80 g      95 t      5 others
ORIGIN
|||||
Query Match      2.5%; Score 70; DB 9; Length 346;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1986 CCTGTAAATCCAGCTACTTGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAGGAGGTGG 2045
|||||
Db 125 CCGTGTAAATCCAGCTACTTGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAGGAGGTGG 66
|||||
QY 2046 AGGTTGCAGT 2055
|||||
Db 65 AGGTTGCAGT 56

RESULT 48
LOCUS AV647070 346 bp mRNA linear EST 15-JAN-2002
DEFINITION AV647070 GLC Homo sapiens cDNA clone GLCAUD05 3', mRNA sequence.
ACCESSION AV647070
VERSION AV647070.1 GI:9868084
KEYWORDS EST.
SOURCE human.

```

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 346)  
Xu.X., Huang.J., Xu.Z., Qian.B., Zhu.Z., Yan.Q., Cai.T., Zhang.X.,  
Xiao.H., Qu.J., Liu.F., Huang.Q., Cheng.Z., Li.N., Du.J., Hu.W.,  
Shen.K., Lu.G., Fu.G., Zhong.M., Xu.S., Gu.W., Huang.W., Zhao.X.,  
Hu.G., Gu.J., Chen.Z. and Han.Z.  
Insight into hepatocellular carcinogenesis at transcriptome level  
by comparing gene expression profiles of hepatocellular carcinoma  
with those of corresponding noncancerous liver  
Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)  
21625106  
Contact: Zeguang Han  
Chinese National Human Genome Center at Shanghai  
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai  
201203, P. R. China  
Tel: 86-21-50801919(ex.45)  
Fax: 86-21-50801922  
Email: hanzg@hgc.sh.cn  
This clone is available at CHGC in Shanghai.  
FEATURES  
source  
1..346  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="GLCAUD05"  
/clone\_lib="GLC"  
/tissue\_type="corresponding non cancerous liver tissue"  
/dev\_stages="Adult"  
/lab\_host="SOLR"  
/note="Vector: pBluescript sk(-); Site\_1: EcoRI; Site\_2:  
XhoI"  
BASE COUNT 89 a 77 c 106 g 74 t  
ORIGIN  
Query Match 2.5%; Score 70; DB 10; Length 346;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1986 CCTGTAATCCAGCTACTTGGAGCGTGGAGGAGAGTGCCTTGAACCCAGGAGGTGG 2045  
|||||  
Db 205 CCTGTAATCCAGCTACTTGGAGCGTGGAGGAGAGTGCCTTGAACCCAGGAGGTGG 264  
QY 2046 AGGTTGCAGT 2055  
|||||  
Db 265 AGGTTGCAGT 274  
RESULT 49  
AQ056722 437 bp DNA linear GSS 30-JUL-1998  
LOCUS  
DEFINITION  
CIT-HSP-233917.TF CIT-HSP Homo sapiens genomic clone 233917, DNA  
sequence.  
ACCESSION AQ056722  
VERSION AQ056722.1 GI:3353328  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 437)  
Adams.M.D., Rounsley.S.D., Zhao.S., Field.C.E., Bass.S., Linher.K.,  
Golden.K., Berry.K., Granger.D., Suh.E., Wible.C., Shizuya.H.,  
Simon.M. and Venter.J.C.  
Use of a random BAC End Sequence Database for Sequence-Ready Map  
Building (1998)  
Unpublished (1998)  
Other\_GSSs: CIT-HSP-233917.TR  
CONTACT: Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200

Fax: 301 838 0208  
Email: mdadams@tigr.org  
Clones are available from Research Genetics (info@resgen.com). BAC  
end search page:  
http://www.tigr.org/tdb/hungen/bac\_end\_search/bac\_end\_search.html.  
Seq primer: M13-21  
Class: BAC ends.  
FEATURES  
Location/Qualifiers  
1..437  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="233917"  
/clone\_lib="CIT-HSP"  
/sex="Male"  
/cell\_type="Sperm"  
/note="Vector: pBelobAC11; Site\_1: HindIII; Site\_2:  
HindIII"  
BASE COUNT 137 a 94 c 118 g 88 t  
ORIGIN  
Query Match 2.5%; Score 70; DB 17; Length 437;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1986 CCTGTAATCCAGCTACTTGGAGCGTGGAGGAGAGTGCCTTGAACCCAGGAGGTGG 2045  
|||||  
Db 282 CCTGTAATCCAGCTACTTGGAGCGTGGAGGAGAGTGCCTTGAACCCAGGAGGTGG 341  
QY 2046 AGGTTGCAGT 2055  
|||||  
Db 342 AGGTTGCAGT 351  
RESULT 50  
N22395/c 448 bp mRNA linear EST 20-DEC-1995  
LOCUS  
DEFINITION  
YW37f08.s1 Morton Fetal Cochlea Homo sapiens cDNA clone  
IMAGE:254439 3' similar to contains Alu repetitive element.; mRNA  
sequence.  
ACCESSION N22395  
VERSION N22395.1 GI:1128529  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 448)  
Hillier,L., Lennon,G., Becker.M., Bonaldo,M.F., Chiapelli,B.,  
Chissoe,S., Dietrich,N., DuBuque,T., Favellio,A., Gish,W., Hawkins  
M., Hultman,M., Kucaba,T., Lacy,M., Le.M., Le.N., Mardis,E., Moore  
B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,  
Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevaskis,E.,  
Underwood,K., Wohlmann,P., Waterston,R., Wilson,R. and Marra,M.  
Generation and analysis of 280,000 human expressed sequence tags  
Genome Res. 6 (9), 807-828 (1996)  
9704478  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
High quality sequence stops: 346  
Source: IMAGE Consortium, LLNL  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 844 Std Error: 0.00  
Seq primer: ml3 -40 forward  
High quality sequence stop: 346.  
FEATURES  
Location/Qualifiers  
1..448  
/organism="Homo sapiens"  
/db\_xref="GDB:3891892"

```
/db_xref="taxon:9606"
/clone="IMAGE:254439"
/clone_lib="Morton Petal Cochlea"
/tissue_type="cochlea"
/lab_stages="16-22 week fetus"
/Note="Organ: ear; Vector: pBluescript SK-; Site_1: EcoRI;
Site_2: XhoI; Reference: Genomics 23, 42-50 (1994) Cloned
unidirectionally. Primer: Oligo dT. Fetal cochlea, normal.
37% of inserts <0.5 Kb, 56% 0.5-1.0 Kb, 7% >1 Kb. Uni-ZAP
XR Vector. Library constructed by N. Robertson, C. Morton.
-5' adaptor sequence: 5' GAATTCGGACGAG 3' -3' adaptor
sequence: 5' CTCGAGTTTTTTTTTTT 3'"
BASE COUNT      124 a      105 c      94 g      122 t      3 others
ORIGIN
```

```
Query Match      2.5%; Score 70; DB 14; Length 448;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1986 CCTGTAATCCCGAGCTACTTTGGAGGCTGAGCGAGGAGTAATCGCTTGAACCCAGGAGGTGG 2045
Db 125 CCTGTAATCCCGAGCTACTTTGGAGGCTGAGCGAGGAGTAATCGCTTGAACCCAGGAGGTGG 66

QY 2046 AGGTTGCAGT 2055
Db 65 AGGTTGCAGT 56
```

```
RESULT 51
AQ284662
LOCUS      AQ284662.1      544 bp      DNA      linear      GSS 04-MAY-1999
DEFINITION      RPC111-7809.TV RPC1-11 Homo sapiens genomic clone RPC1-11-7809, DNA
sequence.
ACCESSION      AQ284662
VERSION      AQ284662.1      GI:3910902
KEYWORDS      GSS.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 544)
Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
Use of human BAC End Sequences for Sequence-Ready Map Building
Unpublished (1998)
Other_GSSs: RPC111-7809.TJ
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are derived from the human BAC library RPC1-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html
Seq primer: 17
Class: BAC ends.
```

```
Location/Qualifiers
1..544
/organism="Homo sapiens"
/db_xref="GDB:7525912"
/db_xref="taxon:9606"
/clone="RPC1-11-7809"
/clone_lib="RPC1-11"
/sex="Male"
/cell_type="Lymphocytes"
/Note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
RPC111 Human Male BAC Library"
```

```
BASE COUNT      178 a      84 c      133 g      149 t
ORIGIN

Query Match      2.5%; Score 70; DB 17; Length 544;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1986 CCTGTAATCCCGAGCTACTTTGGAGGCTGAGCGAGGAGTAATCGCTTGAACCCAGGAGGTGG 2045
Db 439 CCTGTAATCCCGAGCTACTTTGGAGGCTGAGCGAGGAGTAATCGCTTGAACCCAGGAGGTGG 498

QY 2046 AGGTTGCAGT 2055
Db 499 AGGTTGCAGT 508
```

```
RESULT 52
AW973278/c
LOCUS      AW973278      642 bp      mRNA      linear      EST 01-JUN-2000
DEFINITION      EST385376 MAGE resequences, MAGM Homo sapiens cDNA, mRNA sequence.
ACCESSION      AW973278
VERSION      AW973278.1      GI:8163144
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE      1 (bases 1 to 642)
AUTHORS      Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt
,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
Quackenbush,J.
Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray
Unpublished (2000)
```

```
CONTACT: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnq@tigr.org
Plate: 316
Seq primer: Forward.
Location/Qualifiers
1..642
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAGE resequences, MAGM"
/Note="Vector: pBluescriptSKm"
```

```
BASE COUNT      167 a      162 c      130 g      183 t
ORIGIN
```

```
Query Match      2.5%; Score 70; DB 10; Length 642;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1986 CCTGTAATCCCGAGCTACTTTGGAGGCTGAGCGAGGAGTAATCGCTTGAACCCAGGAGGTGG 2045
Db 193 CCTGTAATCCCGAGCTACTTTGGAGGCTGAGCGAGGAGTAATCGCTTGAACCCAGGAGGTGG 134

QY 2046 AGGTTGCAGT 2055
Db 133 AGGTTGCAGT 124
```

```
RESULT 53
AQ741937
LOCUS      AQ741937      649 bp      DNA      linear      GSS 16-JUL-1999
DEFINITION      HS_5569_B2_F07_SP6 RPC1-11 Human Male BAC Library Homo sapiens
genomic clone Plate:1145 Col:14 Row:L, DNA sequence.
ACCESSION      AQ741937
VERSION      AQ741937
KEYWORDS      GSS.
SOURCE      human.
```

```

ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 649)
AUTHORS      Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE        Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL       Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE       99380589
COMMENT      Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University Of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering.bac.htm)
or from Resear h Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 1145 row: L column: 14
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 649.
FEATURES      source
             Location/Qualifiers
             1..649
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone="plate=1145 Col=14 Row=L"
             /clone_lib="RPCI-11 Human Male BAC Library"
             /sex="male"
             /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
             Male blood DNA was isolated from one randomly chosen donor
             and partially digested with a combination of EcoRI and
             EcoRI Methylase. Size selected DNA was cloned into the
             pBACe3.6 vector at EcoRI sites"
BASE COUNT   258 a 88 c 153 g 150 t
ORIGIN
Query Match      2.5%; Score 70; DB 17; Length 649;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1986 CCTGTAATCCACGCTACTTGGGAGGCTGAGCGAGGAGAGATCGCTTGAACCCAGGAGGTGG 2045
|||||
Db 201 CCTGTAATCCACGCTACTTGGGAGGCTGAGCGAGGAGAGATCGCTTGAACCCAGGAGGTGG 260
|||||
QY 2046 AGGTTGCAGT 2055
|||||
Db 261 AGGTTGCAGT 270

RESULT 54
AG185819
LOCUS      AG185819
DEFINITION Pan troglodytes DNA, clone: RP43-060A09.TJ, genomic survey
sequence.
ACCESSION  AG185819
VERSION     AG185819.1 GI:16715499
KEYWORDS    GSS.
SOURCE      Male BAC Library clone:RP43-060A09.TJ.
ORGANISM    Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
AUTHORS     Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE       BAC end sequences of Library RPCI-43

ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     2 (bases 1 to 705)
AUTHORS      Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE       Direct Submission
JOURNAL       Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimpbes@resgen.com, URL:http://hgp.resgen.com.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT      Clones are derived from the chimpanzee BAC library RPCI-43 This BAC
end was generated during the R&D process and may have higher chance
of clone tracking errors.
PRIMERS
Sequencing: TJ
LIBRARY       Vector : pBACe3.6
              R.Site 1 : EcoRI
              R.Site 2 : EcoRI.
              Location/Qualifiers
              1..705
              /organism="Pan troglodytes"
              /db_xref="taxon:9598"
              /clone="RP43-060A09.TJ"
              /sex="male"
              /cell_type="lymphocytes"
              /clone_lib="RPCI-43 Chimpanzee Male BAC Library"
BASE COUNT   240 a 138 c 148 g 175 t
ORIGIN
Query Match      2.5%; Score 70; DB 17; Length 705;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1986 CCTGTAATCCACGCTACTTGGGAGGCTGAGCGAGGAGAGATCGCTTGAACCCAGGAGGTGG 2045
|||||
Db 439 CCTGTAATCCACGCTACTTGGGAGGCTGAGCGAGGAGAGATCGCTTGAACCCAGGAGGTGG 498
|||||
QY 2046 AGGTTGCAGT 2055
|||||
Db 499 AGGTTGCAGT 508

RESULT 55
B14085/c
LOCUS      B14085
DEFINITION A-837A4.TP CIT978SK Homo sapiens genomic clone A-837A4, DNA
sequence.
ACCESSION  B14085
VERSION     B14085.1 GI:2121834
KEYWORDS    GSS.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 571)
AUTHORS     Adams,M.D., Kelley,J.M., Rounsley,S.R. and Venter,J.C.
TITLE       Use of a BAC End Sequence Database for Sequence-Ready Map Building
JOURNAL     Unpublished (1997)
COMMENT      Other_GSSs: 837A4.TV
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tldb/hungen/bac_end_search/bac_end_search.html
Seq primer: SP6
Class: BAC ends.
FEATURES      Location/Qualifiers

```

```
source
1. .571
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="A-837A4"
/clone_lib="CIT978SK"
/sex="Female"
/cell_type="Fibroblast"
/note="Vector: pBAC108L; Site_1: HindIII; Site_2: HindIII;
CalTech Human BAC Library A"
BASE COUNT 116 a 167 c 132 g 156 t
ORIGIN

Query Match 2.4%; Score 69; DB 17; Length 571;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1987 CTGTATCCAGCTACTTGGGAGGCTGAGGAGGAGATCGCTTGAACCCAGGAGTGG 2046
|||||
Db 222 CTGTAATCCAGCTACTTGGGAGGCTGAGGAGGAGATCGCTTGAACCCAGGAGTGG 163
|||||

QY 2047 GGTTCGAGT 2055
|||||
Db 162 GGTTCGAGT 154

RESULT 56
AQ276581
LOCUS
DEFINITION
AQ276581 300 bp DNA linear GSS 22-NOV-1998
CITBI-EI-2521L5.TR CITBI-EI Homo sapiens genomic clone 2521L5, DNA
sequence.
ACCESSION
AQ276581
VERSION
AQ276581.1 GI:3902777
KEYWORDS
GSS.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 300)
AUTHORS
Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
Venter,J.C.
TITLE
Use of a random human BAC End Sequence Database for Sequence-Ready
Map Building
JOURNAL
Unpublished (1998)
COMMENT
Other_GSSs: CITBI-EI-2521L5.TF
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/cdb/hungen/bac_end_search/bac_end_search.html.
Seq primer: M13 Reverse
Class: BAC ends.
FEATURES
Location/Qualifiers
1..300
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2521L5"
/clone_lib="CITBI-EI"
/sex="male"
/cell_type="sperm"
/note="Vector: pBelobAC11; Site_1: EcoRI; Site_2: EcoRI;
CalTech Human BAC Library D"
BASE COUNT 92 a 66 c 84 g 58 t
ORIGIN

Query Match 2.4%; Score 67; DB 17; Length 300;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

source
1..397
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-436N15"
/clone_lib="RPC111 Human Male BAC Library"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBACE3.6; RPC111 Human Male BAC Library"
BASE COUNT 126 a 81 c 121 g 69 t
ORIGIN

Query Match 2.4%; Score 67; DB 17; Length 397;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1986 CCTGTAATCCAGCTACTTGGGAGGCTGAGGAGGAGATCGCTTGAACCCAGGAGTGG 2045
|||||
Db 205 CCTGTAATCCAGCTACTTGGGAGGCTGAGGAGGAGATCGCTTGAACCCAGGAGTGG 264
|||||

QY 2046 AGTTTGC 2052
|||||
Db 265 AGTTTGC 271

RESULT 58
BI023361
LOCUS
DEFINITION
CM4-MT0246-190101-811-d10 MT0246 Homo sapiens cDNA, mRNA sequence.
ACCESSION
BI023361
VERSION
BI023361.1 GI:14429991
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
```



Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 483)  
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM4&t2=CM4-WT0246-190101-811-d10&t3=2001-01-19&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 22  
High quality sequence stop: 470.  
Location/Qualifiers  
1..483  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="WT0246"  
/dev\_stage="Adult"  
/note="Organ: marrow; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 /716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."  
BASE COUNT 158 a 105 c 128 g 92 t  
ORIGIN  
Query Match 2.4%; Score 67; DB 13; Length 483;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1986 CCTGTAATCCAGCTACTTGGAGGCTGAGGAGGAGAGATCGTTGAACCCAGGAGGTGG 2045  
|||||  
Db 387 CCTGTAATCCAGCTACTTGGAGGCTGAGGAGGAGAGATCGTTGAACCCAGGAGGTGG 446  
QY 2046 AGGTTGC 2052  
|||||  
Db 447 AGGTTGC 453  
RESULT 59  
AQ478179  
LOCUS  
DEFINITION  
AQ478179 725 bp DNA linear GSS 23-APR-1999  
RPIC1-11-248K21-TV RPIC1-11 Homo sapiens genomic clone RPIC1-11-248K21  
, DNA sequence.  
ACCESSION  
AQ478179  
VERSION  
AQ478179.1 GI:4660298  
KEYWORDS  
GSS.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 725)  
Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter ,J.C.  
Use of BAC End Sequences from Library RPIC1-11 for Sequence-Ready

Map Building  
Unpublished (1997)  
Other\_GSSs: RPIC1-11-248K21\_TV  
Contact: Shaying Zhao, William Nierman, Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: hbe@tigr.org  
Clones are derived from the human BAC library RPIC1-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genet cs (info@resgen.com). BAC end search page: http://www.tigr.org/tldb/hungen/bac\_end\_search/bac\_end\_search.html.  
Seq primer: T7  
Class: BAC ends.  
Location/Qualifiers  
1..725  
/organism="Homo sapiens"  
/db\_xref="GDB:7595108"  
/db\_xref="taxon:9606"  
/clone\_lib="RPIC1-11-248K21"  
/clone\_lib="RPIC1-11"  
/sex="Male"  
/cell\_type="Lymphocytes"  
/note="Vector: pBACE3.6; Site\_1: EcoRI; Site\_2: EcoRI; RPIC11 Human Male BAC Library"  
BASE COUNT 227 a 157 c 212 g 129 t  
ORIGIN  
Query Match 2.4%; Score 67; DB 17; Length 725;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1986 CCTGTAATCCAGCTACTTGGAGGCTGAGGAGGAGAGATCGTTGAACCCAGGAGGTGG 2045  
|||||  
Db 471 CCTGTAATCCAGCTACTTGGAGGCTGAGGAGGAGAGATCGTTGAACCCAGGAGGTGG 530  
QY 2046 AGGTTGC 2052  
|||||  
Db 531 AGGTTGC 537  
RESULT 60  
AQ750497/c  
LOCUS  
DEFINITION  
AQ750497 841 bp DNA linear GSS 19-JUL-1999  
HS-5576\_B1\_C06\_T7A RPIC1-11 Human Male BAC Library Homo sapiens genomic clone Plate=1152 Col=11 Row=F. DNA sequence.  
ACCESSION  
AQ750497  
VERSION  
AQ750497.1 GI:5537655  
KEYWORDS  
GSS.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 841)  
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.  
Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome  
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)  
99380589  
Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Clones are derived from the human BAC library RPIC1-11. For BAC

library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources ([http://bacpac.med.buffalo.edu/Ordering\\_bac.htm](http://bacpac.med.buffalo.edu/Ordering_bac.htm)) or from Resear h Genetics (info@resgen.com). BAC end Web Server: <http://www.htsc.washington.edu>  
 Plate: 1152 row: F column: 11  
 Seq primer: T7  
 Class: BAC ends  
 High quality sequence stop: 841.

FEATURES  
 source  
 Location/Qualifiers  
 1..841  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="Plate-1152 Col-11 Row=F"  
 /clone\_lib="RPC1-11 Human Male BAC Library"  
 /sex="male"  
 /note="Vector: pBACE3.6; Site\_1: EcoRI; Site\_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at EcoRI sites"  
 BASE COUNT  
 ORIGIN  
 200 a 162 c 186 g 291 t 2 others

Query Match 2.4%; Score 67; DB 17; Length 841;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1986 CCTGTAAATCCAGCTACTTGGAGGCTGAGGAGGAGAGATCGCTTGAACCCAGGAGGTGG 2045  
 |||||||  
 Db 540 CCTGTAAATCCAGCTACTTGGAGGCTGAGGAGGAGAGATCGCTTGAACCCAGGAGGTGG 481

QY 2046 AGTTGTC 2052  
 |||||||  
 Db 480 AGTTGTC 474

RESULT 61  
 AW882256  
 LOCUS  
 DEFINITION  
 RC5-OT0055-080300-021-C12 OT0055 Homo sapiens cDNA, mRNA sequence.  
 ACCESSION  
 AW882256  
 VERSION  
 AW882256.1 GI:8044266  
 EST.  
 KEYWORDS  
 SOURCE  
 human.

ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE  
 1 (bases 1 to 286)  
 AUTHORS  
 Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.  
 TITLE  
 Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL  
 MEDLINE  
 COMMENT  
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 20202663  
 Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome project. This entry can be seen in the following URL (<http://www.ludwig.org.br/scripts/gethtml2.pl?cl=4t2-RC5-OT0055-080300-021-C12&t3=2000-03-08&t4=1>)  
 Seq primer: puc 18 forward  
 High quality sequence stop: 286.

FEATURES  
 source

Location/Qualifiers  
 1..286  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="OT0055"  
 /dev\_stage="Adult"

/note="Organ: ovary; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,776 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."  
 BASE COUNT  
 ORIGIN  
 94 a 62 c 71 g 59 t

Query Match 2.3%; Score 66; DB 10; Length 286;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1986 CCTGTAAATCCAGCTACTTGGAGGCTGAGGAGGAGAGATCGCTTGAACCCAGGAGGTGG 2045  
 |||||||  
 Db 64 CCTGTAAATCCAGCTACTTGGAGGCTGAGGAGGAGAGATCGCTTGAACCCAGGAGGTGG 123

QY 2046 AGTTGTC 2051  
 |||||||  
 Db 124 AGTTGTC 129

RESULT 62  
 AW882261  
 LOCUS  
 DEFINITION  
 RC5-OT0055-080300-021-D10 OT0055 Homo sapiens cDNA, mRNA sequence.  
 ACCESSION  
 AW882261  
 VERSION  
 AW882261.1 GI:8044271  
 EST.  
 KEYWORDS  
 SOURCE  
 human.

ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE  
 1 (bases 1 to 286)  
 AUTHORS

Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.  
 TITLE  
 Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL  
 MEDLINE  
 COMMENT  
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 20202663  
 Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome project. This entry can be seen in the following URL (<http://www.ludwig.org.br/scripts/gethtml2.pl?cl=4t2-RC5-OT0055-080300-021-D10&t3=2000-03-08&t4=1>)  
 Seq primer: puc 18 forward  
 High quality sequence stop: 286.

FEATURES  
 source

Location/Qualifiers  
 1..286  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="OT0055"  
 /dev\_stage="Adult"

/note="Organ: ovary; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived

from ORESTES PCR (U.S. Letters Patent application No. 196  
,716 - Ludwig Institute for Cancer Research) profiles  
into the pUC 18 vector. Reverse transcription of tissue  
mRNA and cDNA amplification were performed under low  
stringency conditions."

BASE COUNT 94 a 62 c 71 g 59 t  
ORIGIN

Query Match 2.3%; Score 66; DB 10; Length 286;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1986 CCTGTAATCCAGCTACTTGGGAGGCTGAGGCAGGAGAAATCGCTTGAACCCAGGAGTGG 2045  
|||||  
Db 64 CCTGTAATCCAGCTACTTGGGAGGCTGAGGCAGGAGAAATCGCTTGAACCCAGGAGTGG 123  
|||||

QY 2046 AGGTTG 2051  
|||||  
Db 124 AGGTTG 129  
|||||

## RESULT 63

AW882265

LOCUS

RC5-OT0055-080300-021-E12 OT0055 Homo sapiens cDNA, mRNA sequence. EST 23-MAY-2000

ACCESSION

AW882265.1 GI:8044275

VERSION

EST.

KEYWORDS

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 286)

Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,

Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,

Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia G.S., Simpson,D.H.,

Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare

,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and

Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2-RC5-OT0055-080

300-021-E12&amp;t3=2000-03-08&amp;t4=1)

Seq primer: puc 18 forward

High quality sequence stop: 286.

Location/Qualifiers

1..286

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="OT0055"

/dev\_stage="Adult"

/note="Organ: ovary; Vector: puc18; Site:1; SmaI; Site:2;

SmaI; A mini-library was made by cloning products derived

from ORESTES PCR (U.S. Letters Patent application No. 196

,716 - Ludwig Institute for Cancer Research) profiles

into the pUC 18 vector. Reverse transcription of tissue

mRNA and cDNA amplification were performed under low

stringency conditions."

BASE COUNT 94 a 62 c 71 g 59 t

ORIGIN

Query Match 2.3%; Score 66; DB 10; Length 286;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1986 CCTGTAATCCAGCTACTTGGGAGGCTGAGGCAGGAGAAATCGCTTGAACCCAGGAGTGG 2045  
|||||  
Db 64 CCTGTAATCCAGCTACTTGGGAGGCTGAGGCAGGAGAAATCGCTTGAACCCAGGAGTGG 123  
|||||

QY 2046 AGGTTG 2051  
|||||  
Db 124 AGGTTG 129  
|||||

## RESULT 64

AW882267

LOCUS

RC5-OT0055-080300-021-F12 OT0055 Homo sapiens cDNA, mRNA sequence. EST 23-MAY-2000

ACCESSION

AW882267.1 GI:8044277

VERSION

EST.

KEYWORDS

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 286)

Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,

Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,

Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia G.S., Simpson,D.H.,

Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare

,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and

Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2-RC5-OT0055-080

300-021-F12&amp;t3=2000-03-08&amp;t4=1)

Seq primer: puc 18 forward

High quality sequence stop: 286.

Location/Qualifiers

1..286

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="OT0055"

/dev\_stage="Adult"

/note="Organ: ovary; Vector: puc18; Site:1; SmaI; Site:2;

SmaI; A mini-library was made by cloning products derived

from ORESTES PCR (U.S. Letters Patent application No. 196

,716 - Ludwig Institute for Cancer Research) profiles

into the pUC 18 vector. Reverse transcription of tissue

mRNA and cDNA amplification were performed under low

stringency conditions."

BASE COUNT 94 a 62 c 71 g 59 t

ORIGIN

QY	2046 AGTTG 2051 
Db	124 AGTTG 129
RESULT 65	
AW518030/c	
LOCUS	xx65h01.x1 NCI_CGAP_Lym12 Homo sapiens cDNA clone IMAGE:2848561 3'
DEFINITION	similar to contains Alu repetitive element,, mRNA sequence.
ACCESSION	AW518030
VERSION	AW518030.1 GI:7156112
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 388)
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL	Unpublished (1997)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs@mail.nih.gov Life Technologies catalog #: 11547-015 DNA sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov/image/html/lresources.shtml Seq primer: -40UP from Gibco High quality sequence stop: 382.
FEATURES	Location/Qualifiers
source	1..388
	/organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:2848561" /clone_lib="NCI_CGAP_Lym12" /tissue_type="lymphoma, follicular mixed small and large cell" /lab_host="DH10B" /note="Organ: lymph node; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.25 kb. Life Technologies catalog #: 11547-015"
BASE COUNT	108 a 107 c 76 g 96 t 1 others
ORIGIN	
Query Match	2.3%; Score 66; DB 10; Length 388;
Best Local Similarity	100.0%; Pred. No. 0;
Matches	66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1990 TAATCCAGCTACTTGGGAGGCTGAGGCAGGAGAATCGCTTGAAACCCAGGAGGTGGAGGT 2049 
Db	128 TAATCCAGCTACTTGGGAGGCTGAGGCAGGAGAATCGCTTGAAACCCAGGAGGTGGAGGT 69 
QY	2050 TGCAGT 2055
Db	68 TGCAGT 63
RESULT 66	
B56190/c	
LOCUS	CIT-HSP-2006D8.TRB CIT-HSP Homo sapiens genomic clone 2006D8, DNA
DEFINITION	sequence.
ACCESSION	B56190
VERSION	B56190.1 GI:2610524
KEYWORDS	GSS.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 395)
AUTHORS	J., Moore B., Schellenberg K., Steptoe M., Tan F., Theising R., White Y., Wyllie T., Waterston R. and Wilson R.
TITLE	WashU-NCI human EST project
JOURNAL	Unpublished (1997)
COMMENT	Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.wustl.edu This clone is available royalty-free through LBNL ; contact the
AUTHORS	Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linlier,K., Golden, .K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M, and Venter,J.C.
TITLE	Use of a random BAC End Sequence Database for Sequence-Ready Map Building
JOURNAL	Unpublished (1997)
COMMENT	Other_GSSs: CIT-HSP-2006D8.TFB Contact: Mark Adams Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: mdadamset@igr.org Clones are available from Research Genetics (info@resgen.com). BAC end search page: http://www.tigr.org/tldb/humgen/bac_end_search/bac_end_search.html Seq primer: M13 Reverse Class: BAC ends.
FEATURES	Location/Qualifiers
source	1..395
	/organism="Homo sapiens" /db_xref="GDB:7039570" /db_xref="taxon:9606" /clone="2006D8" /clone_lib="CIT-HSP" /sex="Male" /cell_type="Sperm" /notes="Vector: pBelobAC11; Site_1: HindIII; Site_2: HindIII"
BASE COUNT	91 a 99 c 77 g 128 t
ORIGIN	
Query Match	2.3%; Score 66; DB 17; Length 395;
Best Local Similarity	100.0%; Pred. No. 0;
Matches	66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1990 TAATCCAGCTACTTGGGAGGCTGAGGCAGGAGAATCGCTTGAAACCCAGGAGGTGGAGGT 204 
Db	171 TAATCCAGCTACTTGGGAGGCTGAGGCAGGAGAATCGCTTGAAACCCAGGAGGTGGAGGT 112 
QY	2050 TGCAGT 2055
Db	111 TGCAGT 106
RESULT 67	
AA679794/c	
LOCUS	ag92b12.s1 Stratagene hNF neuron (#937233) Homo sapiens cDNA clone
DEFINITION	IMAGE:1141919 3' similar to contains Alu repetitive element,, mRNA sequence.
ACCESSION	AA679794
VERSION	AA679794.1 GI:2660316
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 402)
AUTHORS	Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S., Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,R., White,Y., Wyllie,T., Waterston,R. and Wilson,R.
TITLE	WashU-NCI human EST project
JOURNAL	Unpublished (1997)
COMMENT	Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.wustl.edu This clone is available royalty-free through LBNL ; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seq primer: -40ml3 fwd. ET from Amersham  
 High quality sequence stop: 339.

FEATURES  
 source  
 1..402  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1141919"  
 /clone\_lib="Stratagene hNT neuron (#937233)"  
 /dev\_stage="hNT neurons"  
 /lab\_host="SOLR (kanamycin resistant)"  
 /note="vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer: Oligo dT. Differentiated, post mitotic hNT neurons. Average insert size: 1.5 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATCGGCACGAG 3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTT 3'"

BASE COUNT 78 a 108 c 96 g 120 t  
 ORIGIN

Query Match 2.3%; Score 66; DB 9; Length 402;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1990 TAATCCAGCTACTTGGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAGGAGGTGGAGGT 2049  
 Db 274 TAATCCAGCTACTTGGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAGGAGGTGGAGGT 215

Qy 2050 TGCAGT 2055  
 Db 214 TGCAGT 209

RESULT 68  
 B72122/c

LOCUS B72122 414 bp DNA linear GSS 18-JUN-1998  
 DEFINITION CIT978SK-A-699D2.TPC CIT978SK Homo sapiens genomic clone A-699D2,  
 DNA sequence.

ACCESSION B72122  
 VERSION B72122.1 GI:2708894  
 KEYWORDS GSS.

SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 414)  
 AUTHORS Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden  
 K., Berry,K., Granger,D., Suh,E., Wible,C., Kim,U.-J., Shizuya,H.,  
 Simon,M. and Venter,J.C.  
 TITLE Use of a human BAC End Sequence Database for Sequence-Ready Map  
 Building

JOURNAL Unpublished (1997)  
 COMMENT Other\_GSSs: CIT978SK-A-699D2.TVB  
 Contact: Mark Adams  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: mdadams@tigr.org  
 Clones are available from Research Genetics (info@resgen.com). BAC  
 end search page:  
 http://www.tigr.org/tdb/humgen/bac\_end\_search/bac\_end\_search.html  
 Seq primer: SP6  
 Class: BAC ends.

FEATURES  
 source  
 1..414  
 /organism="Homo sapiens"  
 /db\_xref="GDB:1362024"  
 /db\_xref="taxon:9606"  
 /clone="A-699D2"  
 /clone\_lib="CIT978SK"  
 /sex="Female"

/cell\_type="Fibroblast"  
 /note="vector: pBAC108L; Site\_1: HindIII; Site\_2: HindIII;  
 Caltech Human BAC Library A"

BASE COUNT 97 a 102 c 95 g 120 t  
 ORIGIN

Query Match 2.3%; Score 66; DB 17; Length 414;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1986 CCTGTAAATCCAGCTACTTGGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAGGAGGTGG 2045  
 Db 256 CCTGTAAATCCAGCTACTTGGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAGGAGGTGG 197

Qy 2046 AGGTTG 2051  
 Db 196 AGGTTG 191

RESULT 69  
 AQ111323

LOCUS AQ111323 430 bp DNA linear GSS 29-AUG-1998  
 DEFINITION CIT-HSP-2371K16.TR CIT-HSP Homo sapiens genomic clone 2371K16, DNA  
 sequence.

ACCESSION AQ111323  
 VERSION AQ111323.1 GI:3487980  
 KEYWORDS GSS.

SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 430)  
 AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,  
 Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and  
 Venter,J.C.  
 TITLE Use of a random human BAC End Sequence Database for Sequence-Ready  
 Map Building

JOURNAL Unpublished (1998)  
 COMMENT Other\_GSSs: CIT-HSP-2371K16.TF  
 Contact: Mark Adams  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: mdadams@tigr.org  
 Clones are available from Research Genetics (info@resgen.com). BAC  
 end search page:  
 http://www.tigr.org/tdb/humgen/bac\_end\_search/bac\_end\_search.html.  
 Seq primer: M13 Reverse  
 Class: BAC ends.

FEATURES  
 source  
 1..430  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="2371K16"  
 /clone\_lib="CIT-HSP"  
 /sex="Male"  
 /cell\_type="Sperm"  
 /note="vector: pBelobAC11; Site\_1: HindIII; Site\_2:  
 HindIII"

BASE COUNT 124 a 89 c 113 g 103 t 1 others  
 ORIGIN

Query Match 2.3%; Score 66; DB 17; Length 430;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1990 TAATCCAGCTACTTGGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAGGAGGTGGAGGT 2049  
 Db 301 TAATCCAGCTACTTGGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAGGAGGTGGAGGT 360

Qy 2050 TGCAGT 2055



found through the I.M.A.G.E. Consortium/LLNL at:  
image.llnl.gov/image/html/iresources.shtml  
Seq primer: -40UP from Gibco  
High quality sequence stop: 441.

## FEATURES

## source

1..526  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="IMAGE:2907130"  
/clone\_lib="NIH\_MGC\_21"  
/tissue="type="choriocarcinoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: placenta; Vector: pOTB7; Site\_1: XhoI;  
Site\_2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCACGAG(G). Size-selected >500bp  
for average insert size 1.8kb. Library constructed by  
Ling Hong in the laboratory of Gerald M. Rubin (University  
of California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."  
164 a 134 c .96 g 132 t

## BASE COUNT

ORIGIN

Query Match 2.3%; Score 66; DB 10; Length 526;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1990 TAATCCAGCTACTTGGGAGGCTGAGCAGGAGATCGCTTGAACCCAGGAGGTGGAGCT 2049

|||||

Db 129 TAATCCAGCTACTTGGGAGGCTGAGCAGGAGATCGCTTGAACCCAGGAGGTGGAGCT 70

|||||

QY 2050 TGCAGT 2055

|||||

Db 69 TGCAGT 64

## RESULT 73

## BE156416

## LOCUS

BE156416 QVO-HT0368-030100-080-c12 HT0368 Homo sapiens cDNA, mRNA linear EST 21-JUN-2000

BE156416 DEFINITION QVO-HT0368-030100-080-c12 HT0368 Homo sapiens cDNA, mRNA sequence.

BE156416 ACCESSION

BE156416.1 GI:8619137

EST.

KEYWORDS

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 395)

AUTHORS

Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,

Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,

Goldman,G.H., Carvalho,A.F., Matsumura,A., Baia,G.S., Simpson,D.H.,

Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare

,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and

Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

CONTACT: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=&t2=QVO-HT0368-030

100-080-c12&t3=2000-01-03&t4=1)

Seq primer: puc 18 forward

High quality sequence stop: 395.

Location/Qualifiers

## FEATURES

source

## source

1..395  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="HT0368"  
/dev\_stage="Adult"  
/note="Organ: head\_neck; Vector: puc18; Site\_1: SmaI;  
Site\_2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196.716 - Ludwig Institute for Cancer Research)  
profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."  
117 a 79 c 115 g 84 t

## BASE COUNT

ORIGIN

Query Match 2.3%; Score 65; DB 10; Length 395;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1991 AATCCAGCTACTTGGGAGGCTGAGCAGGAGAGTTCGTTGAACCCAGGAGGTGGAGTT 2050

|||||

Db 236 AATCCAGCTACTTGGGAGGCTGAGCAGGAGAGTTCGTTGAACCCAGGAGGTGGAGTT 295

|||||

QY 2051 GCAGT 2055

|||||

Db 296 GCAGT 300

## RESULT 74

## BE156526

## LOCUS

BE156526 QVO-HT0368-310100-091-c07 HT0368 Homo sapiens cDNA, mRNA linear EST 21-JUN-2000

BE156526 DEFINITION QVO-HT0368-310100-091-c07 HT0368 Homo sapiens cDNA, mRNA sequence.

BE156526 ACCESSION

BE156526.1 GI:8619247

EST.

KEYWORDS

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 400)

AUTHORS

Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,

Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,

Goldman,G.H., Carvalho,A.F., Matsumura,A., Baia,G.S., Simpson,D.H.,

Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare

,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and

Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

CONTACT: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=&t2=QVO-HT0368-310

100-091-c07&t3=2000-01-31&t4=1)

Seq primer: puc 18 forward

High quality sequence stop: 9

High quality sequence stop: 400.

Location/Qualifiers

1..400

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="HT0368"

/dev\_stage="Adult"

/note="Organ: head\_neck; Vector: puc18; Site\_1: SmaI;

Site\_2: SmaI; A mini-library was made by cloning products

derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT  
ORIGIN

118 a 80 c 115 g 87 t

Query Match 2.3%; Score 65; DB 10; Length 400;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1991 AATCCAGCTACTTGGAGGCTGAGGAGGAGAGAAATCGCTTGAACCCAGGAGGTGAGGTT 2050

Db 241 AATCCAGCTACTTGGAGGCTGAGGAGGAGAGAAATCGCTTGAACCCAGGAGGTGAGGTT 300

QY 2051 GCAGT 2055

Db 301 GCAGT 305

RESULT 75

BE940056

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

BE940056 443 bp mRNA linear EST 02-OCT-2000  
RC3-UT0034-170800-012-h10 UT0034 Homo sapiens cDNA, mRNA sequence.  
BE940056  
EST.  
GI:10469142

human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 443)  
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,  
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,  
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil

Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=et2-RC3-UT0034-170  
800-012-h10et3-2000-08-17&t4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 176.

Location/Qualifiers  
1..443  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="UT0034"  
/dev\_stage="Adult"

/note="Organ: uterus; tumor: Vector: puc18; Site1: Sma1;  
Site2: Sma1; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

BASE COUNT

ORIGIN

132 a 90 c 134 g 87 t

Query Match

Best Local Similarity

Matches

Conservative

Mismatches

Indels

Gaps

QY 1978

Db 27

QY 2038

Db 87

QY 2038

Db 87

QY 2038

Db 87

QY 2038

Db 87

QY 2038

Db 87

QY 2038

Db 87

QY 2038

Db 87

QY 2038

Db 87

QY 2038

Db 87

QY 2038

Db 87

QY 2038

Db 87

QY 2038

Db 87

QY 2038

Db 87

QY 2038

Db 87

QY 2038

Db 87

QY 2038

Db 87

QY 2038

Db 87

QY 2038

Db 87

QY 2038

Db 87

QY 2038

Db 87

QY 2038

Db 87

QY 2038

Db 87

QY 2038

Db 87

QY 2038

Db 87

QY 2038

Db 87

QY 2038

Db 87

QY 2038

Db 87

QY 2038

Db 87

QY 2038

Db 87

QY 2038

Db 87



BG059756/c  
LOCUS naf53c05.x1 NCI\_CGAP\_Brn65 Homo sapiens cDNA clone IMAGE:4147593 3'  
DEFINITION similar to contains Alu repetitive element;; mRNA sequence.  
ACCESSION BG059756  
VERSION BG059756.1 GI:12527556  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 468)  
AUTHORS NCI/NINDS-CoAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute / National Institute of Neurological Disorders and Stroke, Brain Tumor Genome Anatomy Project (CGAP/BTGP), Tumor Gene Index  
JOURNAL Unpublished (1998)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs@mail.nih.gov](mailto:cgabbs@mail.nih.gov)  
Tissue Procurement: David N. Louis, M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL, send email to: [info@image.llnl.gov](mailto:info@image.llnl.gov)  
Seq primer: -40UP from Gibco  
High quality sequence stop: 416.  
Location/Qualifiers  
1..468  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4147593"  
/clone\_lib="NCI\_CGAP\_Brn65"  
/tissue\_type="glioblastoma without EGFR amplification"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: brain; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dr. Average insert size 1.77 kb. Constructed by Life Technologies."  
BASE COUNT 92 a 135 c 104 g 136 t 1 others  
ORIGIN  
Query Match 2.3%; Score 64; DB 12; Length 468;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1983 GCACCTGTATCCACGCTACTTGGGAGGCTGGAGGAGGAGATCGCTTGAACCCAGGAGG 2042  
Db  
127 GCACCTGTATCCACGCTACTTGGGAGGCTGGAGGAGGAGATCGCTTGAACCCAGGAGG 68  
QY 2043 TGCA 2046  
Db  
67 TGCA 64  
RESULT 78  
AQ344499  
LOCUS AQ344499  
DEFINITION RPCI11-13402.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-13402, DNA sequence.  
ACCESSION AQ344499  
VERSION AQ344499.1 GI:4169395  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 470)  
AUTHORS Zhao, S., Adams, M.D., Nierman, W., Malek, J., de Jong, P. and Venter, J.C.  
TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building

JOURNAL Unpublished (1997)  
COMMENT Contact: Shaying Zhao, William Nierman, Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: [hbe@tigr.org](mailto:hbe@tigr.org)  
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong ([pieter@dejong.med.buffalo.edu](mailto:pieter@dejong.med.buffalo.edu)). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genetics ([info@resgen.com](http://info@resgen.com)). BAC end search page: [http://www.tigr.org/tldb/hungen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tldb/hungen/bac_end_search/bac_end_search.html)  
Seq primer: SP6  
Class: BAC ends.  
FEATURES  
Location/Qualifiers  
1..470  
/organism="Homo sapiens"  
/db\_xref="GDB:7551409"  
/db\_xref="taxon:9606"  
/clone="RPCI-11-13402"  
/clone\_lib="RPCI-11"  
/sex="Male"  
/cell\_type="Lymphocytes"  
/note="Vector: pBACe3.6; Site\_1: EcoRI; Site\_2: EcoRI; RPCI11 Human Male BAC Library"  
BASE COUNT 121 a 100 c 134 g 115 t  
ORIGIN  
Query Match 2.3%; Score 64; DB 17; Length 470;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1979 GCAGCGACCTGTAAATCCAGCTACTTGGAGGCTGGAGGAGGAGATCGCTTGAACCCAG 2038  
Db  
193 GCAGCGACCTGTAAATCCAGCTACTTGGAGGCTGGAGGAGGAGATCGCTTGAACCCAG 252  
QY 2039 GAGG 2042  
Db  
253 GAGG 256  
RESULT 79  
N49425/c  
LOCUS N49425  
DEFINITION YV21c11.r1 Soares fetal liver spleen lNFLS Homo sapiens cDNA clone IMAGE:243380 5' similar to contains Alu repetitive element;; mRNA sequence.  
ACCESSION N49425  
VERSION N49425.1 GI:1190591  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 479)  
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevisakis, E., Waterston, R., Williamson, A., Wohldmann, P. and Willson, R.  
TITLE The WashU-Merck EST Project  
JOURNAL Unpublished (1995)  
COMMENT Contact: Willson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: [est@watson.wustl.edu](mailto:est@watson.wustl.edu)  
This clone is available royalty-free through LLNL; contact the IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Seq primer: T7  
High quality sequence stop: 357.

	FEATURES	SOURCE
1.	100% Pure	100%
2.	100% Pure	100%
3.	100% Pure	100%
4.	100% Pure	100%
5.	100% Pure	100%
6.	100% Pure	100%
7.	100% Pure	100%
8.	100% Pure	100%
9.	100% Pure	100%
10.	100% Pure	100%
11.	100% Pure	100%
12.	100% Pure	100%
13.	100% Pure	100%
14.	100% Pure	100%
15.	100% Pure	100%
16.	100% Pure	100%
17.	100% Pure	100%
18.	100% Pure	100%
19.	100% Pure	100%
20.	100% Pure	100%
21.	100% Pure	100%
22.	100% Pure	100%
23.	100% Pure	100%
24.	100% Pure	100%
25.	100% Pure	100%
26.	100% Pure	100%
27.	100% Pure	100%
28.	100% Pure	100%
29.	100% Pure	100%
30.	100% Pure	100%
31.	100% Pure	100%
32.	100% Pure	100%
33.	100% Pure	100%
34.	100% Pure	100%
35.	100% Pure	100%
36.	100% Pure	100%
37.	100% Pure	100%
38.	100% Pure	100%
39.	100% Pure	100%
40.	100% Pure	100%
41.	100% Pure	100%
42.	100% Pure	100%
43.	100% Pure	100%
44.	100% Pure	100%
45.	100% Pure	100%
46.	100% Pure	100%
47.	100% Pure	100%
48.	100% Pure	100%
49.	100% Pure	100%
50.	100% Pure	100%
51.	100% Pure	100%
52.	100% Pure	100%
53.	100% Pure	100%
54.	100% Pure	100%
55.	100% Pure	100%
56.	100% Pure	100%
57.	100% Pure	100%
58.	100% Pure	100%
59.	100% Pure	100%
60.	100% Pure	100%
61.	100% Pure	100%
62.	100% Pure	100%
63.	100% Pure	100%
64.	100% Pure	100%
65.	100% Pure	100%
66.	100% Pure	100%
67.	100% Pure	100%
68.	100% Pure	100%
69.	100% Pure	100%
70.	100% Pure	100%
71.	100% Pure	100%
72.	100% Pure	100%
73.	100% Pure	100%
74.	100% Pure	100%
75.	100% Pure	100%
76.	100% Pure	100%
77.	100% Pure	100%
78.	100% Pure	100%
79.	100% Pure	100%
80.	100% Pure	100%
81.	100% Pure	100%
82.	100% Pure	100%
83.	100% Pure	100%
84.	100% Pure	100%
85.	100% Pure	100%
86.	100% Pure	100%
87.	100% Pure	100%
88.	100% Pure	100%
89.	100% Pure	100%
90.	100% Pure	100%
91.	100% Pure	100%
92.	100% Pure	100%
93.	100% Pure	100%
94.	100% Pure	100%
95.	100% Pure	100%
96.	100% Pure	100%
97.	100% Pure	100%
98.	100% Pure	100%
99.	100% Pure	100%
100.	100% Pure	100%

```

Qy 1979 GCAGGCACCTGTATCCAGCTACTTGGGAGGCTGAGGAGGAGAGATCGTTGAACCCAG 2038
Db 266 GCAGGCACCTGTATCCAGCTACTTGGGAGGCTGAGGAGGAGAGATCGTTGAACCCAG 207

Qy 2039 GAGG 2042
Db 206 GAGG 203

RESULT 82
LOCUS A2516708 263 bp DNA linear GSS 16-OCT-2000
DEFINITION RPCI-11-195F6.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-195F6,
ACCESSION A2516708
VERSION A2516708
KEYWORDS GSS.
SOURCE A2516708.1 GI:10825382
ORGANISM human.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
,J.C.
TITLE BAC end sequences of library RPCI-11
JOURNAL Unpublished (1997)
COMMENT Other_GSSs: RPCI-11-195F6.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buhalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buhalo.edu/ordering) or from
Research Genet cs (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
This BAC end was generated during the R&D process and may have
higher chance of clone tracking errors.
Seq primer: SP6
Class: BAC ends.
FEATURES
source
1..263
/organism="Homo sapiens"
/db_xref="GDB:7574521"
/db_xref="taxon:9606"
/clone="RPCI-11-195F6"
/clone_lib="RPCI-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
RPCI11 Human Male BAC Library"
BASE COUNT 91 a 60 c 62 g 49 t 1 others
ORIGIN
Query Match 2.2%; Score 63; DB 17; Length 263;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1974 TGGTAGCAGGCACCTGTATCCAGCTACTTGGGAGGCTGAGGAGGAGAGATCGCTTGAA 2033
Db 36 TGGTAGCAGGCACCTGTATCCAGCTACTTGGGAGGCTGAGGAGGAGAGATCGCTTGAA 95

Qy 2034 CCC 2036
Db 96 CCC 98

RESULT 83
BF933309/c
```

```

LOCUS BF933309 380 bp mRNA linear EST 22-JAN-2001
DEFINITION IL5-NT0273-201200-369-e03 NT0273 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF933309
VERSION BF933309.1 GI:12350633
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Negai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jondeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL5&t2=IL5-NT0273-
201200-369-e03&t3=2000-12-20&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 342.
FEATURES
source
1..380
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NT0273"
/dev_stage="Adult"
/note="Organ: nervous tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 77 a 91 c 92 g 119 t 1 others
ORIGIN
Query Match 2.2%; Score 63; DB 12; Length 380;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1993 TCCCAGCTACTTGGGAGGCTGAGCAGGAGAGATCGTTGAACCCAGGAGGTGGC 2052
Db 193 TCCCAGCTACTTGGGAGGCTGAGCAGGAGAGATCGTTGAACCCAGGAGGTGGC 134

Qy 2053 AGT 2055
Db 133 AGT 131

RESULT 84
AI673749/c
```

```

LOCUS AI673749 424 bp mRNA linear EST 15-DEC-1999
DEFINITION tw79ell.x1 NCI_CGAP_Ut3 Homo sapiens cDNA clone IMAGE:2265932 3'
similar to contains Alu repetitive element; contains element THR
repetitive element ;, mRNA sequence.
ACCESSION AI673749
VERSION AI673749.1 GI:4853480
KEYWORDS EST.
SOURCE human.
```

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 424)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emert-Buck, M.D., Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [www-bio.lnl.gov/bbrp/image/image.html](http://www-bio.lnl.gov/bbrp/image/image.html)  
Insert Length: 2630 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 411.  
Location/Qualifiers  
1..424  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone IMAGE:2265932  
/clone\_lib="NCI-CGAP\_Ut3"  
/tissue\_types="poorly-differentiated endometrial adenocarcinoma 2 pooled tumors"  
/lab\_host="DH10B"  
/note="Organ: uterus; Vector: pCMV-SPORT6; Site\_1: SalI; Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.45 kb. Life Technologies catalog #: 11541-018"  
BASE COUNT 93 a 126 c 96 g 109 t  
ORIGIN  
Query Match 2.2%; Score 63; DB 9; Length 424;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1979 GCAGGCACCTGTAATCCAGCTACTTGGAGGCTGAGGAGGAGATCGCTTGAACCCAG 2038  
|||||  
Db 140 GCAGGCACCTGTAATCCAGCTACTTGGAGGCTGAGGAGGAGATCGCTTGAACCCAG 81  
Qy 2039 GAG 2041  
|||||  
Db 80 GAG 78  
RESULT 85  
AQ238522  
LOCUS 1 (bases 1 to 445)  
DEFINITION RPC111-70N2.TJ RPC1-11 Homo sapiens genomic clone RPC1-11-70N2, DNA sequence.  
ACCESSION AQ238522  
VERSION AQ238522.1 GI:3670813  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 445)  
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K., Barry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.  
TITLE Use of human BAC End Sequences for Sequence-Ready Map Building  
JOURNAL Unpublished (1998)  
COMMENT Contact: Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208

Email: mdadams@tigr.org  
Clones are derived from the human BAC library RPC1-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genetics (<http://inforesgen.com>). BAC end search page: [http://www.tigr.org/tdb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html)  
Seq primer: SP6  
Class: BAC ends.  
Location/Qualifiers  
1..445  
/organism="Homo sapiens"  
/db\_xref="GDB:7526809"  
/db\_xref="taxon:9606"  
/clone="RPC1-11-70M2"  
/clone\_lib="RPC1-11"  
/cell\_type="Lymphocytes"  
/notes="Vector: pBACE3.6; Site\_1: EcoRI; Site\_2: EcoRI; RPC111 Human Male BAC Library"  
BASE COUNT 167 a 94 c 94 g 90 t  
ORIGIN  
Query Match 2.2%; Score 63; DB 17; Length 445;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1974 TGTAGCAGGCACCTGTAATCCAGCTACTTGGAGGCTGAGGAGGAGATCGCTTGA 2033  
|||||  
Db 131 TGTAGCAGGCACCTGTAATCCAGCTACTTGGAGGCTGAGGAGGAGATCGCTTGA 190  
Qy 2034 CCC 2036  
|||||  
Db 191 CCC 193  
RESULT 86  
AQ268452  
LOCUS 1 (bases 1 to 454)  
DEFINITION RPC111-69015.TJ RPC1-11 Homo sapiens genomic clone RPC1-11-69015, DNA sequence.  
ACCESSION AQ268452  
VERSION AQ268452.1 GI:3796056  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 454)  
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K., Barry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.  
TITLE Use of human BAC End Sequences for Sequence-Ready Map Building  
JOURNAL Unpublished (1998)  
COMMENT Other\_GSSs: RPC111-69015.TK  
Contact: Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: mdadams@tigr.org  
Clones are derived from the human BAC library RPC1-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genetics (<http://inforesgen.com>). BAC end search page: [http://www.tigr.org/tdb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html)  
Seq primer: SP6  
Class: BAC ends.  
Location/Qualifiers  
1..454  
/organism="Homo sapiens"  
/db\_xref="GDB:7526462"

```

/db_xref="taxon:9606"
/clone="RPCI-11-69015"
/clone_lib="RPCI-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
RPC11 Human Male BAC Library"
BASE COUNT      169 a      98 c      102 g      85 t
ORIGIN

Query Match      2.2%; Score 63; DB 17; Length 454;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1974 TGGTAGCAGGCACCTGTAAATCCCGAGCTACTTGGGAGGCTGAGGAGAGAGATCGCTTGAA 2033
|||||
Db 148 TGGTAGCAGGCACCTGTAAATCCCGAGCTACTTGGGAGGCTGAGGAGAGAGATCGCTTGAA 207
|||||

Qy 2034 CCC 2036
|||
Db 208 CCC 210

RESULT 87
A0017563/c
LOCUS      A0017563      490 bp      DNA      linear      GSS 09-JUN-1998
DEFINITION CIT-HSP-2306J3.TF CIT-HSP Homo sapiens genomic clone 2306J3, DNA
sequence.
ACCESSION  A0017563
VERSION    A0017563.1 GI:3196299
KEYWORDS   GSS.
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS    Adams,M.D., Rounsley,S.D.,Zhao,S., Field,C.E., Bass,S., Linher,K.,
Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
Simon,M. and Venter,J.C.
TITLE      Use of a random BAC End Sequence Database for Sequence-Ready Map
Building (1998)
JOURNAL    Unpublished (1998)
COMMENT    Other_GSSs: CIT-HSP-2306J3.TF
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13-21
Class: BAC ends.
FEATURES             source
     Location/Qualifiers
         1..490
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone="2306J3"
             /clone_lib="CIT-HSP"
             /sex="Male"
             /cell_type="Sperm"
             /note="vector: pBeloBAC11; Site_1: HindIII; Site_2:
HindIII"
BASE COUNT      94 a      135 c      110 g      151 t
ORIGIN

Query Match      2.2%; Score 63; DB 17; Length 490;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1993 TCCGAGCTACTGGGAGGCTGAGCAGGAGAGATCGCTTGAACCCAGGAGGTGGAGGTGC 2052
|||||

```

```

Db 157 TCCGAGCTACTGGGAGGCTGAGCAGGAGAGATCGCTTGAACCCAGGAGGTGGAGGTGC 98
|||||
Qy 2053 AGT 2055
|||
Db 97 AGT 95

RESULT 88
A0624960
LOCUS      A0624960      502 bp      DNA      linear      GSS 16-JUN-1999
DEFINITION CITBI-E1-2649E20.TF CITBI-E1 Homo sapiens genomic clone 2649E20,
DNA sequence.
ACCESSION  A0624960
VERSION    A0624960.1 GI:5087352
KEYWORDS   GSS.
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS    Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and
Venter,J.C.
TITLE      Use of BAC End Sequences from CalTech Libraries for Sequence-Ready
Map Building
JOURNAL    Unpublished (1997)
COMMENT    Other_GSSs: CITBI-E1-2649E20.TR
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbeet@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13-21
Class: BAC ends.
FEATURES             source
     Location/Qualifiers
         1..502
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone="2649E20"
             /clone_lib="CITBI-E1"
             /sex="male"
             /cell_type="Sperm"
             /note="vector: pBeloBAC11; Site_1: EcoRI; Site_2: EcoRI;
Caltech Human BAC Library D"
BASE COUNT      150 a      100 c      123 g      129 t
ORIGIN

Query Match      2.2%; Score 63; DB 17; Length 502;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1993 TCCGAGCTACTGGGAGGCTGAGCAGGAGAGATCGCTTGAACCCAGGAGGTGGAGGTGC 2052
|||||
Db 319 TCCGAGCTACTGGGAGGCTGAGCAGGAGAGATCGCTTGAACCCAGGAGGTGGAGGTGC 378
|||||
Qy 2053 AGT 2055
|||
Db 379 AGT 381

RESULT 89
AG127767
LOCUS      AG127767      639 bp      DNA      linear      GSS 04-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-138116.F, genomic survey sequence.
ACCESSION  AG127767
VERSION    AG127767.1 GI:16656932
KEYWORDS   GSS.
SOURCE     Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male

```

BAC Library clone: PTB-138L16.F.

ORGANISM Pan troglodytes

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pan.

AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.

TITLE BAC end sequences of Library PTB

JOURNAL Unpublished

AUTHORS 2 (bases 1 to 639)

Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.

Direct Submission

TITLE Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: chimbes@sc.riken.go.jp; URL: http://hgp.sc.riken.go.jp/; Tel: 81-45-503-9111, Fax: 81-45-503-9170)

JOURNAL Unpublished

AUTHORS Clones are derived from the chimpanzee BAC library PTB This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.

PRIMERS

Sequencing: -21M13

LIBRARY

Vector : pKS145

R.Site 1 : SacI

R.Site 2 : SacI

Location/Qualifiers

1. 639

/organism="Pan troglodytes"

/db\_xref="taxon:9598"

/clone="PTB-138L16.F"

/sex="male"

/cell\_type="lymphoblast"

/clone.lib="PTB Chimpanzee Male BAC Library"

BASE COUNT 212 a 132 c 141 g 154 t

ORIGIN

Query Match 2.2%; Score 63; DB 17; Length 639;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1993 TCCAGCTACTTGGAGGCTGAGCAGGAGAGTTCGTTGAACCCAGGAGGTGGAGTTGC 2052

|||||

Db 371 TCCAGCTACTTGGAGGCTGAGCAGGAGAGTTCGTTGAACCCAGGAGGTGGAGTTGC 430

QY 2053 AGT 2055

||||

Db 431 AGT 433

RESULT 90

AL704146

LOCUS 683 bp mRNA linear EST 22-MAR-2002

DEFINITION DKFZp68602029\_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone

ACCESSION DKFZp68602029 5', mRNA sequence.

AL704146

VERSION

KEYWORDS

SOURCE EST.

ORGANISM human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

1 (bases 1 to 683)

Blum, H., Bauersachs, S., Mewes, H.W., Gassenhuber, J. and Wiemann, S.

EST (Blum, et al.)

Unpublished (1999)

CONTACT: Blum H

MIPS

Am Klopferspitz 18a D-82152 Martinsried, Germany

This is the 3' sequence of the clone insert

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by LMU (Ludwig Maximilians University, Munich/Germany) within the cDNA sequencing consortium of the German Genome Project.

rl sequence also available.

This clone (DKFZp434K1322) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

Location/Qualifiers

1. 716

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="DKFZp434K1322"

/clone.lib="434 (synonym: htcs3)"

/tissue\_type="testis"

/dev\_stage="adult"

/lab\_host="DH10B"

/note="Vector: pSport1; Site\_1: NotI; Site\_2: SalI"

Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by Oigen (Hilden/Germany) within the cDNA sequencing consortium of the German Genome Project.

No sl sequence available.

This clone (DKFZp686D2029) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

Location/Qualifiers

1. 583

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="DKFZp686D2029"

/clone.lib="686 (synonym: hlcc3)"

/tissue\_type="human skeletal muscle"

/dev\_stage="adult"

/lab\_host="DH10B"

/note="Vector: pTripleX2; Site\_1: SfiIA; Site\_2: SfiIB; cDNA-collection"

#### FEATURES

source

BASE COUNT 191 a 166 c 163 g 163 t

ORIGIN

Query Match 2.2%; Score 63; DB 9; Length 683;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1993 TCCAGCTACTTGGAGGCTGAGCAGGAGAGTTCGTTGAACCCAGGAGGTGGAGTTGC 2052

|||||

Db 254 TCCAGCTACTTGGAGGCTGAGCAGGAGAGTTCGTTGAACCCAGGAGGTGGAGTTGC 313

|||||

QY 2053 AGT 2055

||||

Db 314 AGT 316

RESULT 91

AL042927/c

LOCUS

DEFINITION DKFZp434K1322\_s1 434 (synonym: hlcs3) Homo sapiens cDNA clone

ACCESSION DKFZp434K1322 3', mRNA sequence.

AL042927

VERSION

KEYWORDS

SOURCE EST.

ORGANISM human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

1 (bases 1 to 716)

Blum, H., Bauersachs, S., Mewes, H.W., Gassenhuber, J. and Wiemann, S.

EST (Blum, et al.)

Unpublished (1999)

CONTACT: Blum H

MIPS

Am Klopferspitz 18a D-82152 Martinsried, Germany

This is the 3' sequence of the clone insert

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by LMU (Ludwig Maximilians University, Munich/Germany) within the cDNA sequencing consortium of the German Genome Project.

rl sequence also available.

This clone (DKFZp434K1322) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

Location/Qualifiers

1. 716

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="DKFZp434K1322"

/clone.lib="434 (synonym: htcs3)"

/tissue\_type="testis"

/dev\_stage="adult"

/lab\_host="DH10B"

/note="Vector: pSport1; Site\_1: NotI; Site\_2: SalI"

#### FEATURES

source

```

BASE COUNT      152 a   167 c   141 g   255 t   1 others
ORIGIN

Query Match      2.2%; Score 63; DB 9; Length 716;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1993 TCCAGCTACTTGGGAGGCTGAGGCAGGAGATCGCTTGAACCCAGGAGGTGCGTTC 2052
      |||||||
Db 135 TCCAGCTACTTGGGAGGCTGAGGCAGGAGATCGCTTGAACCCAGGAGGTGCGTTC 76
      |||

QY 2053 AGT 2055
      |||
Db 75 AGT 73

RESULT 92
BQ707295
LOCUS      734 bp      mRNA      linear      EST 16-JUL-2002
DEFINITION AGENCOURT_8475232 NIH_MGC_113 Homo sapiens cDNA clone IMAGE:6301509
5', mRNA sequence.
ACCESSION  BQ707295
VERSION     BQ707295.1 GI:21846194
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 734)
REFERENCE   NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS    National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE      Unpublished (1999)
JOURNAL
COMMENT    Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: Dr. Mark Watson
            cDNA Library Preparation: Rubin Laboratory
            DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
            Clone distribution: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: L1CM2517 row: h column: 22
            High quality sequence stop: 606.
            Location/Qualifiers
            1. .734
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone_lib="NIH_MGC_113"
               /lab_host="DH10B (phage-resistant)"
               /note="Organ: spleen; Vector: pOTB7; Site_1: XhoI; Site_2:
               EcoRI; cDNA made by oligo-dT priming. Directionally cloned
               into EcoRI/XhoI sites using the following 5' adaptor:
               GGCAGGAG(G). Library constructed by Ling Hong in the
               laboratory of Gerald M. Rubin (University of California,
               Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
               Superscript II RT (Life Technologies). Note: this is a
               NIH_MGC Library."
BASE COUNT      227 a   164 c   196 g   146 t   1 others
ORIGIN

Query Match      2.2%; Score 63; DB 14; Length 734;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1979 GCAGGCACCTGTAATCCCACTTGGGAGGCTGAGGAGGAGAGATCGCTTGAACCCAG 2038
      |||||||
Db 586 GCAGGCACCTGTAATCCCACTTGGGAGGCTGAGGAGGAGAGATCGCTTGAACCCAG 645
      |||

QY 2039 GAG 2041
      |||
Db 646 GAG 648

```

```

RESULT 93
BQ755005
LOCUS      1034 bp      mRNA      linear      EST 15-MAY-2001
DEFINITION BQ2711311F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4851819 5',
mRNA sequence.
ACCESSION  BQ755005
VERSION     BQ755005.1 GI:14065658
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 1034)
REFERENCE   NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS    National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE      Unpublished (1999)
JOURNAL
COMMENT    Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: L1CM1695 row: e column: 04
            High quality sequence stop: 622.
            Location/Qualifiers
            1. .1034
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone_lib="NIH_MGC_48"
               /lab_host="DH10B (phage-resistant)"
               /note="Organ: B-cells; Vector: pOTB7; Site_1: XhoI;
               Site_2: EcoRI; cDNA made by oligo-dT priming.
               Directionally cloned into EcoRI/XhoI sites using the
               following 5' adaptor: GGCAGGAG(G). Size-selected >500bp
               for average insert size 1.8Kb. Library constructed by Ling
               Hong in the laboratory of Gerald M. Rubin (University of
               California, Berkeley) using ZAP-cDNA synthesis kit
               (Stratagene) and Superscript II RT (Life Technologies).
               Note: this is a NIH_MGC Library."
BASE COUNT      305 a   211 c   233 g   283 t   2 others
ORIGIN

Query Match      2.2%; Score 62; DB 12; Length 1034;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 919 GGGATACAGAACTGCAGGACCTGGGACCATCTTAAAGTGCACAGTTTCTATGACT 978
      |||||||
Db 457 GGGATACAGAACTGCAGGACCTGGGACCATCTTAAAGTGCACAGTTTCTATGACT 516
      |||

QY 979 TT 980
      ||
Db 517 TT 518

```

```

RESULT 94
AA993041/c
LOCUS      308 bp      mRNA      linear      EST 27-AUG-1998
DEFINITION ct92f06.sl Soares_total_fetus_Nb2HF8.9w Homo sapiens cDNA clone
IMAGE:1624259 3' similar to contains Alu repetitive element; , mRNA
sequence.
ACCESSION  AA993041
VERSION     AA993041.1 GI:3179586
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 308)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 731 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 299.
Location/Qualifiers
1. .308
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1624259"
/clone_lib="Soares_total_fetus_Nb2HF8_9w"
/dev_stage="8-9 weeks"
/lab_host="DH10B"
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from pooled 8-9 week
(total) fetus material with a Not I - oligo(dT) primer [5'
TGTTACCAATCGAAGTGGAGCGCGCTTAATTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 74 a 72 c 82 g 80 t
ORIGIN

Query Match 2.2%; Score 61; DB 9; Length 308;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1982 GGCACCTGTAATCCCGAGCTACTTGGAGGCTGAGGAGAGAAATCGCTTGAACCCAGGAG 2041
|||||
Db 137 GGCACCTGTAATCCCGAGCTACTTGGAGGCTGAGGAGAGAAATCGCTTGAACCCAGGAG 78

QY 2042 G 2042
|
Db 77 G 77

RESULT 95
BM988047/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 396)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. Jose Mercuende
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
The following repetitive elements were found in this cDNA

sequence: 11-296, >ALU (matched complement) 289-351,
>MER5A#DNA/MER1_type
Seq primer: M13 FORWARD
POLYA=Yes.
Location/Qualifiers
1. .396
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE: 3108204"
/clone_lib="NCI_CGAP_DFO"
/tissue_type="Subchondral Bone"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/notes="Organ: Bone; Vector: pT7T3-Pac (Pharmacia) with a
modified polylinker; Site 1: Eco RI; Site 2: Not I;
NCI_CGAP_DFO is a cDNA library containing the following
tissue(s): Subchondral Bone. The library was constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an Eco RI adaptor, digested
with Not I, and cloned directionally into pT7T3-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (GT)18 tail. The
sequence tag for this library is GTTAAGCGTC.
TAG_L1B=UI-H-DFO
TAG_TISSUE=subchondral bone
TAG_SEQ=GTTAAGCGTC"
BASE COUNT 74 a 107 c 83 g 132 t
ORIGIN

Query Match 2.2%; Score 61; DB 14; Length 396;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1982 GGCACCTGTAATCCCGAGCTACTTGGAGGCTGAGGAGAGAAATCGCTTGAACCCAGGAG 2041
|||||
Db 149 GGCACCTGTAATCCCGAGCTACTTGGAGGCTGAGGAGAGAAATCGCTTGAACCCAGGAG 90

QY 2042 G 2042
|
Db 89 G 89

RESULT 96
AA838140/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 400)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David B. Krizman, Ph.D.
DNA Sequencing by: Greg Lennon, Ph.D.
Clone Distribution: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/brp/image/image.html
Insert Length: 1249 Std Error: 0.00

```



Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 353.

## FEATURES

## Source

1. .400  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1385551"  
/clone\_lib="NCI\_CGAP\_Ov2"  
/sex="female"  
/tissue.type="ovary"  
/lab\_host="DH10B"

/note="Vector: pAMP10; mRNA made from invasive ovarian tumor, cDNA made by oligo-dT priming. Non-directionally cloned. Size-selected on agarose gel, average insert size 600 bp. Reference: Krizman et al. (1996) Cancer Research 56:5380-5383."

BASE COUNT 74 a 106 c 102 g 118 t

## ORIGIN

Query Match 2.2%; Score 61; DB 9; Length 400;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1982 GGCACCTGTAAATCCAGCTACTTGGAGGCTGAGGAGAGAGATCGCTTGAACCCAGGAG 2041

Db 136 GGCACCTGTAAATCCAGCTACTTGGAGGCTGAGGAGAGAGATCGCTTGAACCCAGGAG 77

QY 2042 G 2042

Db 76 G 76

## RESULT 97

## AQ596495/c

## LOCUS

HS\_5191\_AL\_DL12\_SPE6 RPCI-11 Human Male BAC Library Homo sapiens

genomic clone Plate=767 Col=23 Row=G, DNA sequence.

## ACCESSION

## AQ596495

## VERSION

## GSS.

## KEYWORDS

## SOURCE

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 404)  
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
Hood,L.

Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome

Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)

99380589

## JOURNAL

## MEDLINE

## COMMENT

Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Clones are derived from the human BAC library RPCI-11. For BAC  
library availability, please contact Pieter de Jong  
(pieter@dejong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering.bac.htm)  
or from Resear h Genetics (info@resgen.com). BAC end Web Server:  
http://www.htsc.washington.edu

Plate: 767 row: G column: 23

Seq primer: SP6

Class: BAC ends

High quality sequence stop: 404.

## FEATURES

## Source

1. .404  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

/clone="Plate=767 Col=23 Row=G"  
/clone\_lib="RPCI-11 Human Male BAC Library"

/sex="male"

/note="Vector: pBACE3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
Male blood DNA was isolated from one randomly chosen donor  
and partially digested with a combination of EcoRI and  
EcoRI Methylase. Size selected DNA was cloned into the  
pBACE3.6 vector at EcoRI sites"

BASE COUNT 104 a 86 c 77 g 137 t

## ORIGIN

Query Match 2.2%; Score 61; DB 17; Length 404;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1995 CCAGCTACTTGGAGGCTGAGGAGGAGAGATCGCTTGAACCCAGGAGTGGAGTTGCAG 2054

Db 359 CCAGCTACTTGGAGGCTGAGGAGGAGAGATCGCTTGAACCCAGGAGTGGAGTTGCAG 300

QY 2055 T 2055

Db 299 T 299

## RESULT 98

## AQ414005/c

## LOCUS

AQ414005 430 bp DNA linear GSS 23-MAR-1999  
RPCI-11-194F22-TV RPCI-11 Homo sapiens genomic clone RPCI-11-194F22

, DNA sequence.

## ACCESSION

## AQ414005

## VERSION

## AQ414005.1

## GI:4473105

## KEYWORDS

## GSS.

## SOURCE

## human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 430)

REFERENCE 1 (bases 1 to 430)

## AUTHORS

## Zhao,S.,

## Adams,M.D.,

## Niernman,W.,

## Malek,J.,

## de Jong,P. and

## Venter

## ,J.C.

## TITLE

## Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready

## Map Building

## Unpublished (1997)

## JOURNAL

## COMMENT

## Other\_GSSs: RPCI-11-194F22.TJ

## Contact: Shaying Zhao, William Niernman, Mark Adams

## Department of Eukaryotic Genomics

## The Institute for Genomic Research

## 9712 Medical Center Dr., Rockville, MD 20850

## Tel: 301 838 0200

## Fax: 301 838 0208

## Email: hbe@tigr.org

## Clones are derived from the human BAC library RPCI-11. For BAC

## library availability, please contact Pieter de Jong

## (pieter@dejong.med.buffalo.edu). Clones may be purchased from

## BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from

## Research Genet cs (info@resgen.com). BAC end search page:

## http://www.tigr.org/tldb/hungen/bac\_end\_search/bac\_end\_search.html.

## Seq primer: T7

## Class: BAC ends.

## FEATURES

## Location/Qualifiers

## 1. .430

## /organism="Homo sapiens"

## /db\_xref="GDB:7574253"

## /db\_xref="taxon:9606"

## /clone="RPCI-11-194F22"

## /clone\_lib="RPCI-11"

## /sex="Male"

## /cell\_type="Lymphocytes"

## /note="Vector: pBACE3.6; Site\_1: EcoRI; Site\_2: EcoRI;

## RPCI11 Human Male BAC Library"

## BASE COUNT 111 a 92 c 86 g 141 t

## ORIGIN

## Query Match 2.2%; Score 61; DB 17; Length 430;

Search completed: July 17, 2003, 23:12:39  
Job time : 3603 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 17, 2003, 21:32:20 ; Search time 7088 seconds  
(without alignments)  
11570.500 Million cell updates/sec

Title: US-09-966-880A-7  
Perfect score: 2818  
Sequence: 1 agagaccatcatttaattga.....aaaaaaaaaaaaaaaaaaaaa 2818

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 500 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_htg.\*

3: gb\_in.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vi.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_mu.\*

20: em\_om.\*

21: em\_or.\*

22: em\_ov.\*

23: em\_pat.\*

24: em\_ph.\*

25: em\_pi.\*

26: em\_ro.\*

27: em\_sts.\*

28: em\_un.\*

29: em\_vi.\*

30: em\_htg\_hum.\*

31: em\_htg\_inv.\*

32: em\_htg\_other.\*

33: em\_htg\_mus.\*

34: em\_htg\_pln.\*

35: em\_htg\_rtd.\*

36: em\_htg\_mam.\*

37: em\_htg\_vrt.\*

38: em\_sy.\*

39: em\_htgo\_hum.\*

40: em\_htgo\_mus.\*

41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2791	99.0	2791	9	AB040431	AB040431 Homo sapi
2	2174	77.1	11204	9	AB040430	AB040430 Homo sapi
3	2174	77.1	71132	9	AC092184	AC092184 Homo sapi
4	1815	64.4	1837	9	BC006296	BC006296 Homo sapi
5	587	20.8	596	12	AF529829	AF529829 Mus muscu
6	584	20.7	596	9	AF529823	AF529823 Homo sapi
7	572	20.3	597	12	AF529846	AF529846 Crictetulu
8	569	20.2	597	12	AF529842	AF529842 Crictetulu
9	548	19.4	597	12	AF529847	AF529847 Crictetulu
10	546	19.4	597	12	AF529828	AF529828 Mus muscu
11	545	19.3	596	9	AF529815	AF529815 Homo sapi
12	545	19.3	596	9	AF529816	AF529816 Homo sapi
13	545	19.3	596	9	AF529819	AF529819 Homo sapi
14	545	19.3	596	9	AF529820	AF529820 Homo sapi
15	545	19.3	596	9	AF529821	AF529821 Homo sapi
16	545	19.3	596	9	AF529822	AF529822 Homo sapi
17	545	19.3	596	9	AF529824	AF529824 Homo sapi
18	545	19.3	596	9	AF529825	AF529825 Homo sapi
19	545	19.3	596	9	AF529826	AF529826 Homo sapi
20	545	19.3	596	9	AF529827	AF529827 Homo sapi
21	545	19.3	596	12	AF529830	AF529830 Mus muscu
22	545	19.3	596	12	AF529831	AF529831 Mus muscu
23	545	19.3	596	12	AF529833	AF529833 Mus muscu
24	545	19.3	596	12	AF529835	AF529835 Mus muscu
25	545	19.3	596	12	AF529837	AF529837 Mus muscu
26	545	19.3	596	12	AF529839	AF529839 Mus muscu
27	544	19.3	595	12	AF529834	AF529834 Mus muscu
28	535	19.0	597	12	AF529841	AF529841 Crictetulu
29	535	19.0	597	12	AF529843	AF529843 Crictetulu
30	535	19.0	597	12	AF529844	AF529844 Crictetulu
31	535	19.0	597	12	AF529845	AF529845 Crictetulu
32	535	19.0	597	12	AF529848	AF529848 Crictetulu
33	535	19.0	597	12	AF529849	AF529849 Crictetulu
34	535	19.0	597	12	AF529850	AF529850 Crictetulu
35	535	19.0	597	12	AF529851	AF529851 Crictetulu
36	535	19.0	597	12	AF529852	AF529852 Crictetulu
37	535	19.0	597	12	AF529853	AF529853 Crictetulu
38	535	19.0	597	12	AF529854	AF529854 Crictetulu
39	529	18.8	591	12	AF529855	AF529855 Crictetulu
40	526	18.7	577	12	AF529840	AF529840 Mus muscu
41	516	18.3	596	9	AF529818	AF529818 Homo sapi
42	494	17.5	596	9	AF529817	AF529817 Homo sapi
43	494	17.5	596	12	AF529832	AF529832 Mus muscu
44	494	17.5	596	12	AF529836	AF529836 Mus muscu
45	494	17.5	596	12	AF529838	AF529838 Mus muscu
46	487	17.3	547	12	AF529856	AF529856 Crictetulu
47	86	3.1	203250	9	AC007240	AC007240 Homo sapi
48	79	2.8	156540	9	AC079329	AC079329 Homo sapi
49	79	2.8	187836	2	AC023634	AC023634 Homo sapi
50	77	2.7	41511	9	AC004603	AC004603 Homo sapi
51	77	2.7	103287	2	AC084034	AC084034 Homo sapi
52	77	2.7	106018	9	HS864118	AC031233 Human DNA
53	77	2.7	109538	9	AC091854	AC091854 Homo sapi
54	77	2.7	121028	9	AY052369	AY052369 Homo sapi
55	77	2.7	126295	9	AC079944	AC079944 Homo sapi
56	77	2.7	154968	2	AC078928	AC078928 Homo sapi
57	77	2.7	162958	2	AC011022	AC011022 Homo sapi
58	77	2.7	177777	9	AC007342	AC007342 Homo sapi
59	77	2.7	181183	2	AL391866	AL391866 Homo sapi
60	77	2.7	186431	2	AC022281	AC022281 Homo sapi
61	77	2.7	188863	9	AC007345	AC007345 Homo sapi
62	77	2.7	192539	9	AP005264	AP005264 Homo sapi
63	77	2.7	194871	9	CNS01BRV	AL118538 Human chr
64	77	2.7	225432	9	AF027390	AF027390 Homo sapi

66	c	76	2.7	153300	2	AC011933	AC011933 Homo sapi	139	70	2.5	2460	9	AK095243	Homo sapi
67	c	76	2.7	199725	2	AC100787	Homo sapi	140	70	2.5	22996	9	AC087175	Homo sapi
68	c	74	2.6	66452	2	AC024328	Homo sapi	c 141	70	2.5	23379	9	AL356755	Human DNA
69	c	70	2.6	66792	9	AC008754	Homo sapi	c 142	70	2.5	23574	6	AX326810	Sequence
70	c	74	2.6	118958	9	AC092579	Homo sapi	c 143	70	2.5	23574	9	AF466288	Homo sapi
71	c	74	2.6	158213	2	AF214635	Homo sapi	c 144	70	2.5	30972	9	AC123763	Homo sapi
72	c	74	2.6	158605	9	AC105081	Homo sapi	c 145	70	2.5	35651	9	AL353798	Homo sapi
73	c	74	2.6	160371	2	AC068793	Homo sapi	c 146	70	2.5	36921	9	AL353798	Homo sapi
74	c	74	2.6	162200	9	AC012156	Homo sapi	c 147	70	2.5	37490	9	AC004185	Homo sapi
75	c	74	2.6	167722	9	AC073548	Homo sapi	c 148	70	2.5	38209	9	AC026356	Homo sapi
76	c	74	2.6	169234	9	AC018633	Homo sapi	c 149	70	2.5	40709	9	AC005542	Homo sapi
77	c	74	2.6	172816	9	AC093899	Homo sapi	c 150	70	2.5	41716	9	AC011522	Structure o
78	c	74	2.6	176355	9	AC025518	Homo sapi	c 151	70	2.5	43657	9	AC097506	Homo sapi
79	c	74	2.6	176932	9	AC016772	Homo sapi	c 152	70	2.5	44336	9	AC010525	Homo sapi
80	c	74	2.6	182909	9	AC026130	Homo sapi	c 153	70	2.5	62117	2	AC107891	Homo sapi
81	c	74	2.6	186870	2	AC116170	Homo sapi	c 154	70	2.5	63621	9	AL590867	Homo sapi
82	c	74	2.6	187697	2	AC063929	Homo sapi	c 155	70	2.5	64864	9	AC025260	Human DNA
83	c	74	2.6	201886	9	AC092700	Homo sapi	c 156	70	2.5	66395	9	HS119823	Human DNA
84	c	74	2.6	205952	9	AC024940	Homo sapi	c 157	70	2.5	67550	2	AC016452	Human DNA s
85	c	74	2.6	228652	9	CNS01DWD	Human chr	c 158	70	2.5	67763	2	AC103752	Human DNA
86	c	74	2.6	250681	9	AP004282	Human chr	c 159	70	2.5	67895	9	AL138807	Human DNA
87	c	73	2.6	79319	9	AL512274	Human DNA	c 160	70	2.5	68420	2	AC130285	Human DNA
88	c	73	2.6	96165	9	AC009702	Human DNA	c 161	70	2.5	68468	9	HUMMBDBC	Human DNA f
89	c	73	2.6	141175	2	AC012110	Homo sapi	c 162	70	2.5	74549	9	HSJ890015	Human DNA
90	c	73	2.6	144355	2	AC016743	Homo sapi	c 163	70	2.5	84132	9	AC005348	Human DNA
91	c	73	2.6	153378	2	AC023536	Homo sapi	c 164	70	2.5	85643	2	AC015787	Homo sapi
92	c	73	2.6	167116	9	AC009305	Homo sapi	c 165	70	2.5	86894	2	AL390319	Homo sapi
93	c	73	2.6	176176	2	AC079360	Homo sapi	c 166	70	2.5	97621	9	AL607089	Homo sapi
94	c	73	2.6	179470	2	AL162211	Homo sapi	c 167	70	2.5	97845	9	AL355337	Human DNA
95	c	73	2.6	183101	9	AC021915	Human DNA	c 168	70	2.5	100000	9	AP000509	Human DNA
96	c	73	2.6	191652	9	AL354668	Human DNA	c 169	70	2.5	106868	2	AC003023	Homo sapi
97	c	73	2.6	202950	2	AC017100	Human DNA	c 170	70	2.5	110000	2	AC080008	Homo sapi
98	c	73	2.6	216441	2	AP001336	Human DNA	c 171	70	2.5	110000	2	AC080008	Homo sapi
99	c	73	2.6	239008	2	AC022460	Human DNA	c 172	70	2.5	110000	2	Continuation (2 of	Continuation (2 of
100	c	72	2.6	89211	9	AL669821	Human DNA	c 173	70	2.5	110000	2	Continuation (3 of	Continuation (3 of
101	c	72	2.6	104417	9	AL672220	Human DNA	c 174	70	2.5	110000	2	Continuation (4 of	Continuation (4 of
102	c	72	2.6	115596	9	AL672220	Human DNA	c 175	70	2.5	110000	2	Continuation (2 of	Continuation (2 of
103	c	72	2.6	129272	9	AL451125	Human DNA	c 176	70	2.5	110144	9	AC008053	Homo sapi
104	c	72	2.6	143697	2	AL358572	Human DNA	c 177	70	2.5	111848	2	AC124612	Homo sapi
105	c	72	2.6	149827	2	AL390880	Homo sapi	c 178	70	2.5	113687	9	AC069281	Homo sapi
106	c	72	2.6	154244	2	AC073463	Homo sapi	c 179	70	2.5	116236	9	AL354877	Homo sapi
107	c	72	2.6	156503	9	AC011511	Homo sapi	c 180	70	2.5	117752	9	AC008205	Homo sapi
108	c	72	2.6	160903	2	AP002776	Homo sapi	c 181	70	2.5	118847	9	AC013413	Homo sapi
109	c	72	2.6	161267	9	AC008378	Homo sapi	c 182	70	2.5	121856	9	AC022575	Homo sapi
110	c	72	2.6	162526	2	AC024182	Human DNA	c 183	70	2.5	123280	2	AC025662	Homo sapi
111	c	72	2.6	162973	9	AL138958	Human DNA	c 184	70	2.5	123395	9	HS889N15	Human DNA
112	c	72	2.6	163795	9	AP000356	Human DNA	c 185	70	2.5	124102	9	AC021148	Human DNA
113	c	72	2.6	165165	9	AC095031	Homo sapi	c 186	70	2.5	126312	9	AC000026	Homo sapi
114	c	72	2.6	167386	2	AL159993	Homo sapi	c 187	70	2.5	128024	9	AL590455	Human sapi
115	c	72	2.6	171566	2	AP001885	Homo sapi	c 188	70	2.5	129413	9	HSJ709L21	Human sapi
116	c	72	2.6	173785	9	AP000355	Homo sapi	c 189	70	2.5	132202	9	AC025218	Human sapi
117	c	72	2.6	183690	9	AC091180	Homo sapi	c 190	70	2.5	132573	9	AC016564	Human sapi
118	c	72	2.6	202943	9	AC010134	Human sapi	c 191	70	2.5	135886	9	HSX1ST2	Human Xq1
119	c	72	2.6	209519	9	AC008397	Human sapi	c 192	70	2.5	140606	2	AL353761	Homo sapi
120	c	71	2.5	92975	9	AC002369	Homo sapi	c 193	70	2.5	140919	2	AC034181	Homo sapi
121	c	71	2.5	110116	9	AP001931	Homo sapi	c 194	70	2.5	142439	9	AC008749	Homo sapi
122	c	71	2.5	130642	2	AP001082	Homo sapi	c 195	70	2.5	145206	9	AC026170	Homo sapi
123	c	71	2.5	144588	2	AC012239	Homo sapi	c 196	70	2.5	146360	9	HS1156N12	Human sapi
124	c	71	2.5	152210	9	AC007953	Homo sapi	c 197	70	2.5	148179	9	AP001631	Human sapi
125	c	71	2.5	153394	2	AP000727	Homo sapi	c 198	70	2.5	149266	2	AC021562	Homo sapi
126	c	71	2.5	158349	2	AC009438	Homo sapi	c 199	70	2.5	149438	2	AC010264	Homo sapi
127	c	71	2.5	161298	2	AP002858	Homo sapi	c 200	70	2.5	149870	9	AC024579	Homo sapi
128	c	71	2.5	161397	9	AC090218	Homo sapi	c 201	70	2.5	149913	2	AC010330	Homo sapi
129	c	71	2.5	163390	9	AL137058	Human DNA	c 202	70	2.5	150355	9	HSJ360H10	Human sapi
130	c	71	2.5	165139	9	AC103549	Human sapi	c 203	70	2.5	152129	9	HSJ360M22	Human sapi
131	c	71	2.5	170540	2	AC092707	Homo sapi	c 204	70	2.5	153215	2	HS24A17	Human sapi
132	c	71	2.5	180949	9	AL365495	Human sapi	c 205	70	2.5	156395	2	AC023014	Homo sapi
133	c	71	2.5	186325	2	AC074244	Human sapi	c 206	70	2.5	156422	9	AC093579	Homo sapi
134	c	71	2.5	193281	2	AC079376	Homo sapi	c 207	70	2.5	157193	9	AL355315	Human sapi
135	c	71	2.5	194296	9	AC095351	Homo sapi	c 208	70	2.5	157385	9	AC006023	Human sapi
136	c	71	2.5	195101	9	AC110615	Homo sapi	c 209	70	2.5	157827	9	AC027260	Homo sapi
137	c	71	2.5	199992	2	AC021522	Homo sapi	c 210	70	2.5	157963	9	AP002986	Homo sapi
138	c	71	2.5	236723	2	AC068771	Homo sapi	c 211	70	2.5	158135	2	AC080040	Homo sapi

C 212	70	2.5	158302	2	AC069481	Homo sapi	C 285	70	2.5	194219	2	AC023054	Homo sapi
C 213	70	2.5	159298	2	AL450994	Homo sapi	C 286	70	2.5	194771	2	AC109912	Homo sapi
C 214	70	2.5	159849	2	AC021454	Homo sapi	C 287	70	2.5	195310	2	AC024189	Homo sapi
C 215	70	2.5	160659	2	AC036235	Homo sapi	C 288	70	2.5	196806	9	AC104212	Homo sapi
C 216	70	2.5	161442	2	AL683875	Homo sapi	C 289	70	2.5	197837	9	AC093107	Homo sapi
C 217	70	2.5	161572	9	AC106897	Homo sapi	C 290	70	2.5	198694	9	AC012671	Homo sapi
C 218	70	2.5	161835	9	AC098484	Homo sapi	C 291	70	2.5	201155	9	AC009086	Homo sapi
C 219	70	2.5	162133	2	AC068569	Homo sapi	C 292	70	2.5	201508	2	AC026290	Homo sapi
C 220	70	2.5	162343	2	AC109319	Homo sapi	C 293	70	2.5	201707	2	AC131309	Homo sapi
C 221	70	2.5	163915	2	AC087451	Homo sapi	C 294	70	2.5	202414	2	AC090966	Papio cyn
C 222	70	2.5	164302	2	AC127457	Homo sapi	C 295	70	2.5	204037	2	AC015881	Homo sapi
C 223	70	2.5	164352	2	AC024008	Homo sapi	C 296	70	2.5	204298	9	AL662844	Human DNA
C 224	70	2.5	164429	9	AC026366	Homo sapi	C 297	70	2.5	206329	2	AC017113	Homo sapi
C 225	70	2.5	164499	9	AC092418	Homo sapi	C 298	70	2.5	206647	9	AP002898	Homo sapi
C 226	70	2.5	164650	2	AC093016	Homo sapi	C 299	70	2.5	207433	6	AX326809	Sequence
C 227	70	2.5	164838	2	AL691479	Homo sapi	C 300	70	2.5	208334	2	AC009730	Homo sapi
C 228	70	2.5	165447	9	AC027575	Homo sapi	C 301	70	2.5	209287	9	AC012077	Homo sapi
C 229	70	2.5	165499	2	AC131238	Homo sapi	C 302	70	2.5	220480	2	AC023973	Homo sapi
C 230	70	2.5	165861	2	AC106812	Homo sapi	C 303	70	2.5	228219	2	AC128648	Homo sapi
C 231	70	2.5	166867	9	AP003733	Homo sapi	C 304	70	2.5	230375	2	AC004085	Homo sapi
C 232	70	2.5	167343	6	AX332231	Sequence	C 305	70	2.5	236822	9	DB4394	Homo sapi
C 233	70	2.5	167343	6	AX335067	Sequence	C 306	70	2.5	260409	9	AC004019	Homo sapi
C 234	70	2.5	167343	9	HSU96629	Human chrom	C 307	70	2.5	286758	9	AC006449	Homo sapi
C 235	70	2.5	167357	2	AC055771	Homo sapi	C 308	70	2.5	306464	2	AC068708	Homo sapi
C 236	70	2.5	167631	2	AC016032	Homo sapi	C 309	70	2.5	321974	2	AC090682	Homo sapi
C 237	70	2.5	167668	9	AL353804	Human DNA	C 310	70	2.5	323263	2	AC079406	Homo sapi
C 238	70	2.5	168043	2	AC011935	Homo sapi	C 311	70	2.5	329643	2	AL512381	Homo sapi
C 239	70	2.5	168366	9	AC007546	Homo sapi	C 312	70	2.5	340000	9	AP001748	Homo sapi
C 240	70	2.5	168431	2	AC040972	Homo sapi	C 313	69	2.4	440003	9	AF317635	Homo sapi
C 241	70	2.5	168495	2	AC084299	Homo sapi	C 314	69	2.4	49996	9	HSJ816K9	Human DNA
C 242	70	2.5	168880	9	AC017104	Homo sapi	C 315	69	2.4	83120	9	HS796117	Human DNA
C 243	70	2.5	168887	9	AC106796	Homo sapi	C 316	69	2.4	88210	9	AL607062	Human DNA
C 244	70	2.5	168887	9	AL662833	Human DNA	C 317	69	2.4	126736	2	AC044788	Homo sapi
C 245	70	2.5	168964	2	AC131212	Homo sapi	C 318	69	2.4	129517	2	AC016400	Homo sapi
C 246	70	2.5	169032	9	AC016931	Homo sapi	C 319	69	2.4	147999	2	AL161635	Homo sapi
C 247	70	2.5	169118	2	AC096645	Homo sapi	C 320	69	2.4	148295	9	AC090510	Homo sapi
C 248	70	2.5	170444	2	AL359977	Homo sapi	C 321	69	2.4	154037	9	AC072054	Homo sapi
C 249	70	2.5	170569	2	AC108109	Homo sapi	C 322	69	2.4	160815	2	AP001895	Homo sapi
C 250	70	2.5	170882	2	AC023429	Homo sapi	C 323	69	2.4	162191	9	AC109458	Homo sapi
C 251	70	2.5	170970	9	AL354707	Human DNA	C 324	69	2.4	163372	2	AC087744	Homo sapi
C 252	70	2.5	171829	9	AC022116	Homo sapi	C 325	69	2.4	163569	2	AC023293	Homo sapi
C 253	70	2.5	172049	2	AC008414	Homo sapi	C 326	69	2.4	165373	9	AC067957	Homo sapi
C 254	70	2.5	173029	2	AC002059	Homo sapi	C 327	69	2.4	172525	9	AC008514	Homo sapi
C 255	70	2.5	173085	2	AC127524	Homo sapi	C 328	69	2.4	185075	2	AC118583	Papio cyn
C 256	70	2.5	174058	2	AC068695	Homo sapi	C 329	69	2.4	186755	9	AC024022	Homo sapi
C 257	70	2.5	174231	9	AC021054	Homo sapi	C 330	69	2.4	186783	9	AC090355	Homo sapi
C 258	70	2.5	175072	2	AC022358	Homo sapi	C 331	69	2.4	186902	9	AC010863	Homo sapi
C 259	70	2.5	175625	2	AC068475	Homo sapi	C 332	69	2.4	187167	2	AP001563	Homo sapi
C 260	70	2.5	175980	2	AC025859	Homo sapi	C 333	69	2.4	191436	9	AL158830	Human DNA
C 261	70	2.5	176206	9	AC009967	Homo sapi	C 334	69	2.4	228652	9	CNS01DWD	Human chr
C 262	70	2.5	179441	2	AC044879	Homo sapi	C 335	69	2.4	52350	9	AC092330	Homo sapi
C 263	70	2.5	179843	2	AC131159	Homo sapi	C 336	69	2.4	69901	9	HS694E4	Human DNA
C 264	70	2.5	180392	9	AC025945	Homo sapi	C 337	68	2.4	87402	9	HSJ437M21	Human DNA
C 265	70	2.5	181923	2	AC022872	Homo sapi	C 338	68	2.4	94568	9	AC114973	Homo sapi
C 266	70	2.5	182408	2	AL358772	Homo sapi	C 339	68	2.4	100072	9	AL137144	Human DNA
C 267	70	2.5	184268	2	AC011401	Homo sapi	C 340	68	2.4	108893	9	AC079169	Homo sapi
C 268	70	2.5	184362	2	AL450268	Homo sapi	C 341	68	2.4	114023	2	HS316B5	Human sapien
C 269	70	2.5	184470	9	AL589823	Human DNA	C 342	68	2.4	125055	9	HS167F1	Human DNA
C 270	70	2.5	184825	9	AL391380	Human DNA	C 343	68	2.4	126152	9	HS433M19	s
C 271	70	2.5	185574	9	AC073341	Homo sapi	C 344	68	2.4	129338	9	AC016597	Homo sapi
C 272	70	2.5	186135	2	AC124947	Homo sapi	C 345	68	2.4	130824	9	AC078880	Homo sapi
C 273	70	2.5	186558	2	AC079031	Homo sapi	C 346	68	2.4	158143	2	AC016893	Homo sapi
C 274	70	2.5	186780	9	AC005740	Homo sapi	C 347	68	2.4	162124	2	AC113191	Homo sapi
C 275	70	2.5	188049	2	AC069391	Homo sapi	C 348	68	2.4	164076	9	AC021059	Homo sapi
C 276	70	2.5	188141	2	AC023120	Homo sapi	C 349	68	2.4	165287	2	AC034165	Homo sapi
C 277	70	2.5	188481	9	AL138836	Human DNA	C 350	68	2.4	175670	2	AC103953	Homo sapi
C 278	70	2.5	188645	9	AC012047	Homo sapi	C 351	68	2.4	178336	9	AC073359	Homo sapi
C 279	70	2.5	188823	2	AC016890	Homo sapi	C 352	68	2.4	185551	9	HSJ17K16	Human DNA
C 280	70	2.5	188846	9	AC009030	Homo sapi	C 353	68	2.4	187469	2	AC058793	Homo sapi
C 281	70	2.5	189156	9	AC008669	Homo sapi	C 354	68	2.4	188948	9	AC005832	Homo sapi
C 282	70	2.5	189972	2	AC107214	Homo sapi	C 355	68	2.4	197652	9	AC012074	Homo sapi
C 283	70	2.5	191872	2	AC129837	Papio cyn	C 356	68	2.4	201973	2	AC087314	Homo sapi
C 284	70	2.5	194197	9	AC064871	Homo sapi	C 357	68	2.4	224044	2	AC010882	Homo sapi

C 358	67	2.4	363	11	GL3005	GL13005 SWSS2845 Er
C 359	67	2.4	401	9	HSU41198	U41198 Human (TTTG
C 360	67	2.4	4565	9	AF187983	Homo sapi
C 361	67	2.4	5055	2	AC005843	Homo sapi
C 362	67	2.4	30960	2	AC003102	Homo sapi
C 363	67	2.4	34164	2	AC090913	Homo sapi
C 364	67	2.4	51407	9	AL357556	Human DNA
C 365	67	2.4	52459	9	AF228703	Homo sapi
C 366	67	2.4	61294	9	AL671934	Human DNA
C 367	67	2.4	64167	9	HSBA8982	Human DNA
C 368	67	2.4	67131	2	AC103791	Homo sapi
C 369	67	2.4	82700	2	AC015635	Homo sapi
C 370	67	2.4	82700	2	AC015635	Homo sapi
C 371	67	2.4	93385	2	AC109513	Homo sapi
C 372	67	2.4	96002	2	AC109326	Homo sapi
C 373	67	2.4	101077	2	AC024051	Homo sapi
C 374	67	2.4	104228	9	AL354751	Human DNA
C 375	67	2.4	118234	9	AL355476	Human DNA
C 376	67	2.4	124001	9	HS886K2	Human DNA
C 377	67	2.4	124814	9	AL161436	Human DNA
C 378	67	2.4	137830	9	AC003029	Human DNA
C 379	67	2.4	142060	2	AL353642	Homo sapi
C 380	67	2.4	147888	2	AC016538	Homo sapi
C 381	67	2.4	149266	2	AC116022	Homo sapi
C 382	67	2.4	153875	9	AC003682	Homo sapi
C 383	67	2.4	156997	2	AC011465	Homo sapi
C 384	67	2.4	159322	9	AC026085	Homo sapi
C 385	67	2.4	161788	9	AC009314	Homo sapi
C 386	67	2.4	162728	9	AP002906	Homo sapi
C 387	67	2.4	164034	9	AC011489	Homo sapi
C 388	67	2.4	167072	2	AC060780	Homo sapi
C 389	67	2.4	170695	9	AC016334	Homo sapi
C 390	67	2.4	174582	9	HUAC003010	Homo sapi
C 391	67	2.4	176121	2	AC013489	Homo sapi
C 392	67	2.4	176709	9	AC011737	Homo sapi
C 393	67	2.4	176973	2	AC130449	Homo sapi
C 394	67	2.4	180774	2	AC061979	Homo sapi
C 395	67	2.4	181213	2	AC002399	Homo sapi
C 396	67	2.4	183599	2	AC073501	Homo sapi
C 397	67	2.4	188846	9	AC009030	Homo sapi
C 398	67	2.4	189271	9	AL138752	Human DNA
C 399	67	2.4	195648	2	AF215848	Homo sapi
C 400	67	2.4	199451	2	AF215848	Homo sapi
C 401	67	2.4	202638	2	AC103959	Homo sapi
C 402	67	2.4	206596	2	AC087624	Homo sapi
C 403	67	2.4	208612	2	AC108668	Homo sapi
C 404	67	2.4	214530	9	AC022098	Homo sapi
C 405	67	2.4	217931	9	AC022098	Homo sapi
C 406	67	2.4	242970	2	AC004803	Homo sapi
C 407	67	2.4	288888	9	HSB310932	Homo sapi
C 408	67	2.4	298575	2	HSBA2284	Homo sapi
C 409	66	2.3	2405	9	AK094131	Homo sapi
C 410	66	2.3	2847	9	AK094891	Homo sapi
C 411	66	2.3	34639	9	AC005765	Homo sapi
C 412	66	2.3	53196	9	AC108047	Homo sapi
C 413	66	2.3	62449	2	AC061985	Homo sapi
C 414	66	2.3	78036	9	AC005755	Homo sapi
C 415	66	2.3	89250	2	AC016334	Homo sapi
C 416	66	2.3	90476	9	AC015910	Homo sapi
C 417	66	2.3	92975	9	AC002369	Homo sapi
C 418	66	2.3	109445	9	AC008946	Homo sapi
C 419	66	2.3	111107	9	AC005924	Homo sapi
C 420	66	2.3	111682	9	AC005162	Homo sapi
C 421	66	2.3	112573	9	AL353151	Human DNA
C 422	66	2.3	123149	9	AC008655	Homo sapi
C 423	66	2.3	124310	2	AC008587	Homo sapi
C 424	66	2.3	124949	9	AL358794	Human DNA
C 425	66	2.3	130609	2	AC023217	Homo sapi
C 426	66	2.3	130853	9	AC091974	Homo sapi
C 427	66	2.3	132644	9	AC104801	Homo sapi
C 428	66	2.3	134462	9	AC001231	Genomic s
C 429	66	2.3	142698	9	AC025429	Homo sapi
C 430	66	2.3	147140	9	AC010226	Homo sapi

C 431	66	2.3	147201	2	AC053469	Homo sapi
C 432	66	2.3	148046	9	AC034226	Homo sapi
C 433	66	2.3	148911	2	AL353142	Homo sapi
C 434	66	2.3	149125	2	AC069401	Homo sapi
C 435	66	2.3	150791	9	AL162458	Human DNA
C 436	66	2.3	151600	2	AC010278	Homo sapi
C 437	66	2.3	152709	2	AC114799	Homo sapi
C 438	66	2.3	152794	9	AL359085	Human DNA
C 439	66	2.3	155067	9	AC010301	Homo sapi
C 440	66	2.3	155470	9	AC104448	Homo sapi
C 441	66	2.3	157377	2	AC022827	Homo sapi
C 442	66	2.3	159007	9	AC020658	Homo sapi
C 443	66	2.3	160161	2	AC007846	Homo sapi
C 444	66	2.3	160417	2	AC024317	Homo sapi
C 445	66	2.3	163540	2	AC024229	Homo sapi
C 446	66	2.3	163197	2	AC091392	Homo sapi
C 447	66	2.3	164500	2	AC069171	Homo sapi
C 448	66	2.3	165496	9	AC104461	Homo sapi
C 449	66	2.3	167254	9	CNS05TDS	Human chr
C 450	66	2.3	167390	2	AC007263	Homo sapi
C 451	66	2.3	168751	2	AC068969	Homo sapi
C 452	66	2.3	168956	2	AC090286	Homo sapi
C 453	66	2.3	170883	9	AC067805	Homo sapi
C 454	66	2.3	171452	2	AC002118	Homo sapi
C 455	66	2.3	171452	2	AP000476	Homo sapi
C 456	66	2.3	172654	2	AC067737	Homo sapi
C 457	66	2.3	172867	2	AC026207	Homo sapi
C 458	66	2.3	173226	9	CNS01DTQ	Human chr
C 459	66	2.3	175652	9	AC084864	Homo sapi
C 460	66	2.3	176715	2	AC073474	Homo sapi
C 461	66	2.3	176861	9	AC097065	Homo sapi
C 462	66	2.3	177819	9	AC100771	Homo sapi
C 463	66	2.3	179462	2	AC021102	Homo sapi
C 464	66	2.3	180736	2	AC023925	Homo sapi
C 465	66	2.3	182288	9	AC018362	Homo sapi
C 466	66	2.3	188050	2	AL391845	Homo sapi
C 467	66	2.3	190619	9	AC108172	Homo sapi
C 468	66	2.3	191426	2	AC026086	Homo sapi
C 469	66	2.3	193281	2	AC079376	Homo sapi
C 470	66	2.3	194545	2	AC087482	Homo sapi
C 471	66	2.3	198017	9	AC012314	Homo sapi
C 472	66	2.3	198316	2	AC019157	Homo sapi
C 473	66	2.3	200430	9	AC011500	Homo sapi
C 474	66	2.3	205307	2	CNS01RG3	Homo sapi
C 475	66	2.3	205307	2	AC009968	Homo sapi
C 476	66	2.3	208396	9	AC003101	Homo sapi
C 477	66	2.3	209752	2	AC079620	Homo sapi
C 478	66	2.3	212030	2	AC019148	Homo sapi
C 479	66	2.3	213865	9	AC093166	Homo sapi
C 480	66	2.3	214017	2	AC026359	Homo sapi
C 481	66	2.3	214162	9	AC084865	Homo sapi
C 482	66	2.3	220807	9	AC093334	Homo sapi
C 483	66	2.3	340000	9	AP001690	Homo sapi
C 484	65	2.3	2687	9	AK026456	Homo sapi
C 485	65	2.3	28136	9	HSL184D6	Human DNA s
C 486	65	2.3	67187	9	AC108744	Homo sapi
C 487	65	2.3	98387	9	AC003077	Human IAC
C 488	65	2.3	102477	9	AC008935	Homo sapi
C 489	65	2.3	107919	2	AL391274	Homo sapi
C 490	65	2.3	137496	9	AC005288	Homo sapi
C 491	65	2.3	146868	2	AC016802	Homo sapi
C 492	65	2.3	148120	9	AC090051	Homo sapi
C 493	65	2.3	153148	2	AC017030	Homo sapi
C 494	65	2.3	154968	9	AL596342	Human DNA
C 495	65	2.3	160615	9	AC112236	Homo sapi
C 496	65	2.3	160903	2	AP002776	Homo sapi
C 497	65	2.3	162294	2	AC102321	Homo sapi
C 498	65	2.3	166558	2	AC112141	Homo sapi
C 499	65	2.3	169339	2	AC110008	Homo sapi
C 500	65	2.3	169892	9	AC068397	Homo sapi

```

RESULT_1
LOCUS       AB040431                2791 bp      mRNA          linear     PRI 03-OCT-2000
DEFINITION   Homo sapiens AID mRNA for activation-induced cytidine deaminase, complete CDS.
ACCESSION    AB040431
VERSION      AB040431.1 GI:9988409
KEYWORDS     AID; activation-induced cytidine deaminase; Human AID.
SOURCE       Homo sapiens cDNA to mrna..
ORGANISM     Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (sites)
AUTHORS      Muto,T., Muramatsu,M., Taniwaki,M., Kinoshita,K. and Honjo,T.
TITLE        Isolation, tissue distribution, and chromosomal localization of the human activation-induced cytidine deaminase (AID) gene
JOURNAL      Genomics 68 (1), 85-88 (2000)
MEDLINE      PubMed ID# 10780890
REFERENCE    2 (sites)
AUTHORS      Revy,P., Muto,T., Levy,Y., Geismann,F., Plehani,A., Sanal,O., Catalan,N., Forgeville,M., Dufourcq-Lagelouse,R., Gennery,A., Tezcan,I., Ersoy,F., Kayserili,H., Ugazio,A.G., Brousse,N., Muramatsu,M., Notarangelo,L.D., Kinoshita.K., Honjo.T., Fischer.A. and Durandy.A.
TITLE        Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2)
JOURNAL      Cell 102 (5), 565-575 (2000)
MEDLINE      PubMed ID# 10780541
REFERENCE    3 (bases 1 to 2791)
AUTHORS      Muto,T., Muramatsu,M., Taniwaki.M., Kinoshita,K. and Honjo.T.
TITLE        Direct Submission
JOURNAL      Submitted (18-MAR-2000) Tasuku Honjo, Kyoto University, Department of Medical Chemistry, Faculty of Medicine; Yoshida, Sakyo-ku, Kyoto, Kyoto 606-8501, Japan (E-mail:honjo@four.med.kyoto-u.ac.jp, Tel.:81-75-753-4371(ex.4371), Fax:81-75-753-4388)
FEATURES             Location/Qualifiers
     source           1..2791
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene              1..2791
                     /gene="AID"
     CDS               77..673
                     /gene="AID"
                     /codon_start=1
                     /product="activation-induced cytidine deaminase"
                     /protein_id="BAB12721.1"
                     /db_xref="GI:9988410"
                     /translation="MSDLNRRKFLYQFNWAKGRRFTLYCYVVKRSDATSFSLDFGLRNKGCHVELLRLFIVISDWLDPGRYCVTWTSPSCDARHVADPLRGFKNLSLRIPTALFYCEDRKAPESGRILRHAGVOIAIMTFKDIFYCWNITFFVENHERIFPKAWGLHNVSRLSRQLLRILLPLYVEVDLDAFRFTGL"
BASE COUNT         842 a 548 g 625 q 776 t
ORIGIN
Query Match                      99.0%; Score 2791.; DB 9; Length 2791;
Best Local Similarity 100.0%; Pred.No. 0;
Matches 2791; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      4 GAACCATCAATTAAATGAAGTGAGACTTTTCCTGGCGCTGAGACTTGCAGGGAGGCCAAGAAGAA 63
| | | | | 
Db      1 GAACCATCAATTAAATGAAGTGAGACTTTTCCTGGCGCTGAGACTTCCAGGGAGGCCAAGAAG 60
| | | | | 

Qy      64 CACTCTGGCACCACTATGTGACAGCCCTCTGTGATGACCGGAGGAAGTTTTCTTTACCAAATT 123
| | | | | 
Db      61 CACTCTGGCACCACTATGTGACAGCCCTCTGTGATGACCGGAGGAAGTTTTCTTTACCAAATT 120
| | | | | 

Qy      124 CAAAAATGTCGGCTGGCGTAAGGCTGGCGCTGAGACCTACCTGCTGCTACCTAGTAGGAGAG 183
| | | | | 
Db      121 CAAAATGTCGGCTGGCGTAAGGCTGGCGCTGAGACCTACCTGCTGCTAGTAGTAGGAGAG 180
| | | | | 

Qy      184 CGCTGACAGTGCACATPCTCTTTTCACTPGGACTTTTGTTATCTTCGCAATAAGAACGGCTG 243

```

Db 1261 AAGCATGTTTATGTTGTACAAAAGAAGATTGTTATGGTGGGATGGAGGTATAGAC 1320  
QY 1324 CATGATGTCACCTTCAAGCTACTTTAATAAGGATCTTAAATGGGCGAGGACACTGT 1383  
Db 1321 CATGATGTCACCTTCAAGCTACTTTAATAAGGATCTTAAATGGGCGAGGACACTGT 1380  
QY 1384 GAACAGACACCTTAATAATGGGTGTAGTGTGAAGTAGCAAAATCTTCTGGAAGCGCAA 1443  
Db 1381 GAACAGACACCTTAATAATGGGTGTAGTGTGAAGTAGCAAAATCTTCTGGAAGCGCAA 1440  
QY 1444 CTCCTTTAAGGAAGTCCCTAATTTAGAACACCCACAACTTCACATATCATTAATAGCA 1503  
Db 1441 CTCCTTTAAGGAAGTCCCTAATTTAGAACACCCACAACTTCACATATCATTAATAGCA 1500  
QY 1504 AACAAATGGAAGAAAGTTGCTTGAATGTTGGGAGAGGAAATCTATTGGCTCTCGTGGG 1563  
Db 1501 AACAAATGGAAGAAAGTTGCTTGAATGTTGGGAGAGGAAATCTATTGGCTCTCGTGGG 1560  
QY 1564 TCTCTTCATCTCAGAAATGCCAATCAGGTCGAAGTTTGCTACATTTTGATGTGTGAT 1623  
Db 1561 TCTCTTCATCTCAGAAATGCCAATCAGGTCGAAGTTTGCTACATTTTGATGTGTGAT 1620  
QY 1624 GCTTCTCCCAAGGTATATTAACTATATAAGAGAGTTGTGACAAACAGAAATGATAAGC 1683  
Db 1621 GCTTCTCCCAAGGTATATTAACTATATAAGAGAGTTGTGACAAACAGAAATGATAAGC 1680  
QY 1684 TCGAACCGTGGCACACGCTCATAGTTCTAGCTGCTGGGAGTTGAGAGGAGGATGG 1743  
Db 1681 TCGAACCGTGGCACACGCTCATAGTTCTAGCTGCTGGGAGTTGAGAGGAGGATGG 1740  
QY 1744 CTTGAACACAGTGTTCAGGCCAGCCTGGGCAACATACAAAGATCCTGTCTCTCAAAA 1803  
Db 1741 CTTGAACACAGTGTTCAGGCCAGCCTGGGCAACATACAAAGATCCTGTCTCTCAAAA 1800  
QY 1804 AAAAAAAGGAGAGAGAGGCGCGGCTGTGCTCAGCGCTCTAATCCAG 1863  
Db 1801 AAAAAAAGGAGAGAGAGGCGCGGCTGTGCTCAGCGCTCTAATCCAG 1860  
QY 1864 CACTTTGGAGCGGAGCGCGGAGATCACTGTGGTCAGAGTTTGAGACGACCTGGC 1923  
Db 1861 CACTTTGGAGCGGAGCGCGGAGATCACTGTGGTCAGAGTTTGAGACGACCTGGC 1920  
QY 1924 CAACATGCAAAACCCCTGTCTACTCAAAATGCAAAATTAGCCAGCGTGTAGCAGG 1983  
Db 1921 CAACATGCAAAACCCCTGTCTACTCAAAATGCAAAATTAGCCAGCGTGTAGCAGG 1980  
QY 1984 CACCTGTAATCCAGCTACTTTGGAGGCTGAGGAGGAGATCGCTTGAACCCAGGAGT 2043  
Db 1981 CACCTGTAATCCAGCTACTTTGGAGGCTGAGGAGGAGATCGCTTGAACCCAGGAGT 2040  
QY 2044 GGAGTTGCAGTAAGCTGAGATCGTGGCTGACCTCCAGCTGGGCGACAGAGCAAGA 2103  
Db 2041 GGAGTTGCAGTAAGCTGAGATCGTGGCTGACCTCCAGCTGGGCGACAGAGCAAGA 2100  
QY 2104 CTCGTCTCAGAAAAAAG 2163  
Db 2101 CTCGTCTCAGAAAAAAG 2160  
QY 2164 GAGAGGATGGGAAGCATTCAGAGAAATGTCCTTTATCCAAACAAATGTAAGGAGCC 2223  
Db 2161 GAGAGGATGGGAAGCATTCAGAGAAATGTCCTTTATCCAAACAAATGTAAGGAGCC 2220  
QY 2224 AATAAGGATCCCTATTGTCTCTTTTGGTGTCTATTGTCCTTAACTGCTTTGAC 2283  
Db 2221 AATAAGGATCCCTATTGTCTCTTTTGGTGTCTATTGTCCTTAACTGCTTTGAC 2280  
QY 2284 AGTGAGAAAAATATTCAGAAATACCATATCCCTGTGCGGTATTATACCTAGCAACCTTGC 2343  
Db 2281 AGTGAGAAAAATATTCAGAAATACCATATCCCTGTGCGGTATTATACCTAGCAACCTTGC 2340  
QY 2344 AATCAAGATGAGCAGATCCACAGAGAAACTTGAATGCACAACTGCTTTAATCTTA 2403  
Db 2341 AATGAAGATGAGCAGATCCACAGAGAAACTTGAATGCACAACTGCTTTAATCTTA 2400

QY 2404 TTGTACATAAGCTTTCTAAAAAGAGTTAAAAATTTGTTACTTTCATGTTATTCATTATATTTTAA 2463  
Db 2401 TTGTACATAAGCTTTCTAAAAAGAGTTAAAAATTTGTTACTTTCATGTTATTCATTATATTTTAA 2460  
QY 2464 TATTATTTTGGCTCTAATGATTTTATTAAACATGATTCTCTTTTCTGATATATTGAAAT 2523  
Db 2461 TATTATTTTGGCTCTAATGATTTTATTAAACATGATTCTCTTTTCTGATATATTGAAAT 2520  
QY 2524 GGAGTCTCAAAAGCTTCATAAATTTATAACTTTTAAAGATGATTCTTAATAAACACCTATCTA 2583  
Db 2521 GGAGTCTCAAAAGCTTCATAAATTTATAACTTTTAAAGATGATTCTTAATAAACACCTATCTA 2580  
QY 2584 ATTCTTAACATTTGCAGTAAATGCTGCTACGAAGCCATTTCTCTTGGATTTTACTAAACCTTTT 2643  
Db 2581 ATTCTTAACATTTGCAGTAAATGCTGCTACGAAGCCATTTCTCTTGGATTTTACTAAACCTTTT 2640  
QY 2644 ATGACAGCAAAATTTGCTTCTGGCTCACCTTTCAATCAGTTAAATAAATGATAAATTAATTTT 2703  
Db 2641 ATGACAGCAAAATTTGCTTCTGGCTCACCTTTCAATCAGTTAAATAAATGATAAATTAATTTT 2700  
QY 2704 GGAAGCTGTGAAGATAAAATACCAAAATAAATAATATAAAGTGAATTTATATGAAGTTAA 2763  
Db 2701 GGAAGCTGTGAAGATAAAATACCAAAATAAATAATATAAAGTGAATTTATATGAAGTTAA 2760  
QY 2764 AATAAAAAATCAGTATGATGAATAAACTTGG 2794  
Db 2761 AATAAAAAATCAGTATGATGAATAAACTTGG 2791  
RESULT 2  
AB040430  
LOCUS Homo sapiens AID gene for activation-induced cytidine deaminase, complete cds.  
DEFINITION  
AB040430  
ACCESSION  
VERSION AB040430.1 GI:9988407  
KEYWORDS AID: activation-induced cytidine deaminase.  
SOURCE Homo sapiens DNA.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE  
AUTHORS Muto, T., Muramatsu, M., Taniwaki, M., Kinoshita, K. and Honjo, T.  
TITLE Isolation, tissue distribution, and chromosomal localization of the human activation-induced cytidine deaminase (AID) gene  
JOURNAL Genomics 68 (1), 85-88 (2000)  
MEDLINE 20408890  
REFERENCE  
AUTHORS Revy, P., Muto, T., Levy, Y., Geissmann, F., Plebani, A., Sanjal, O., Catalan, N., Forveille, M., Boursier, R., Gennery, A., Tezcan, I., Ersoy, F., Kayserli, H., Ugazio, A.G., Brousse, N., Muramatsu, M., Notarangelo, L.D., Kinoshita, K., Honjo, T., Fischer, A. and Durandy, A.  
TITLE Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2)  
JOURNAL Cell 102 (5), 565-575 (2000)  
MEDLINE 20460541  
REFERENCE  
AUTHORS Muto, T., Muramatsu, M., Taniwaki, M., Kinoshita, K. and Honjo, T.  
TITLE Direct Submission  
JOURNAL Submitted (18-MAR-2000) Tasuku Honjo, Kyoto University, Department of Medical Chemistry, Faculty of Medicine, Yoshida, Sakyo-ku, Kyoto, Kyoto 606-8501, Japan (E-mail: honjo@four.med.kyoto-u.ac.jp, Tel: 81-75-753-4371(ex.4371), Fax: 81-75-753-4388)  
FEATURES  
source 1. 11204  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
gene join(521..528,6280..6427,7807..8077,8371..8486,8956..9009)  
/gene="AID"  
CDS join(521..528,6280..6427,7807..8077,8371..8486,8956..9009)  
/gene="AID"



/codon_start=1 /product="activation-induced cytidine deaminase" /protein_id="Bab12720.1" /db_xref="GI:9988408" /translation="MDSLMMNRKFLYQFNVPWAKGRRETYLYVYVVRDSTSL DFGLRNKGCHVELLFRLTISWDLDPRCYRWTFSWSPCYDCARHVADFLRNP NUSLRFARLYFCDKAEPEGLRLHRAGVOIAIMTKDFYCWNTVEVNHRTFK AWGLHNSVRLSRORLRLLPLYEYDDLDAFRTGL" BASE COUNT 3305 a 2273 c 2373 g 3253 t ORIGIN									
Query Match 77.1%; Score 2174; DB 9; Length 11204; Best Local Similarity 100.0%; Pred. No. 0; Matches 2174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
Qy	622	GCCCTGTATGAGGTTGATGACTTACGAGACGCAATTCGTACTTTGGGACTTTGTATAGCA	681						
Db	8955	GCCCTGTATGAGGTTGATGACTTACGAGACGCAATTCGTACTTTGGGACTTTGTATAGCA	9014						
Qy	682	ACTTCCAGGAATGTACACACAGATGAATAATCTCTGCTGAAGACAGTGGATAAAAAACAG	741						
Db	9015	ACTTCCAGGAATGTACACACAGATGAATAATCTCTGCTGAAGACAGTGGATAAAAAACAG	9074						
Qy	742	TCCTTCAAGTCCTCTGTTTTATCTTCAACTCTCACTTCTTAGAGTTTACAGAAA	801						
Db	9075	TCCTTCAAGTCCTCTGTTTTATCTTCAACTCTCACTTCTTAGAGTTTACAGAAA	9134						
Qy	802	AATATTTATATAGCACTCTTTAAAAAGATCTATGCTTTGAAAAATAGAGAAGAACACAGG	861						
Db	9135	AATATTTATATAGCACTCTTTAAAAAGATCTATGCTTTGAAAAATAGAGAAGAACACAGG	9194						
Qy	862	TCGTGGCAGGAGCGTGTGCAATTTGGTGAGTTTGAATGCAACATTTGCCCTACTGGG	921						
Db	9195	TCGTGGCAGGAGCGTGTGCAATTTGGTGAGTTTGAATGCAACATTTGCCCTACTGGG	9254						
Qy	922	AATAACAGAACTGCAGGACCTGGAGCATCTTAAAGTGTCAACGTTTCTATGACTTTT	981						
Db	9255	AATAACAGAACTGCAGGACCTGGAGCATCTTAAAGTGTCAACGTTTCTATGACTTTT	9314						
Qy	982	AGGTAGGATGAGACGAGAAGGTAGATCTCTAAAAAGCATGGTGAGAGATCAAAATGTTTT	1041						
Db	9315	AGGTAGGATGAGACGAGAAGGTAGATCTCTAAAAAGCATGGTGAGAGATCAAAATGTTTT	9374						
Qy	1042	ATATCAACATCCCTTTATTTGATTTGATTTGAGTTTAACAGTGGTGTAGTGATGATTT	1101						
Db	9375	ATATCAACATCCCTTTATTTGATTTGATTTGAGTTTAACAGTGGTGTAGTGATGATTT	9434						
Qy	1102	TTCTATTCTTTTCCCTTGACGTTTACTTTTCAAGTAAACAAACTCTTCCATCAGGCCATG	1161						
Db	9435	TTCTATTCTTTTCCCTTGACGTTTACTTTTCAAGTAAACAAACTCTTCCATCAGGCCATG	9494						
Qy	1162	ATCTATAGGACCTCCTAATGAGAGTATCTGGGTGATTTGACCCCAAAACCATCTCTCCAA	1221						
Db	9495	ATCTATAGGACCTCCTAATGAGAGTATCTGGGTGATTTGACCCCAAAACCATCTCTCCAA	9554						
Qy	1222	AGCATTAATATCAACATCGCTGTATGTTTAAATCAGCAGAGCATGTTTATGTTT	1281						
Db	9555	AGCATTAATATCAACATCGCTGTATGTTTAAATCAGCAGAGCATGTTTATGTTT	9614						
Qy	1282	GTACAAAAGAGATTTGTTATGGGTGGGATGGAGGTATAGACCATGCGTCACCTTCA	1341						
Db	9615	GTACAAAAGAGATTTGTTATGGGTGGGATGGAGGTATAGACCATGCGTCACCTTCA	9674						
Qy	1342	AGCTACTTTTAAAGGATCTTAAATGGCAGGAGGACTGTGAACAAGACACCCCTAATA	1401						
Db	9675	AGCTACTTTTAAAGGATCTTAAATGGCAGGAGGACTGTGAACAAGACACCCCTAATA	9734						
Qy	1402	ATGGTGTGATGCTGAAGTAGCAAACTCTCTGAAAGCGAACTCTTTTAAAGGAGTCCC	1461						
Db	9735	ATGGTGTGATGCTGAAGTAGCAAACTCTCTGAAAGCGAACTCTTTTAAAGGAGTCCC	9794						
Qy	1462	TAATTTAGAAACACCCCAACACTTCACATATCATATAATAGCAACAATTCGAGGAGATT	1521						

Db	9795	TAATTTAGAAACACCCCAACACTTCACATATCATATAATAGCAACAATTTGGAAGGAGTT	9854						
Qy	1522	GCTTTGAATGTTGGGAGAGAGAAAATCTATTGGGCTCTCGTGGTCTCTTTCATCTCAGAAAT	1581						
Db	9855	GCTTTGAATGTTGGGAGAGAGAAAATCTATTGGGCTCTCGTGGTCTCTTTCATCTCAGAAAT	9914						
Qy	1582	GCCAATCAGGTCAAGGTTTGCTACATTTTGTATGTGTGTGATGCTTCTCCCAAGGTATA	1641						
Db	9915	GCCAATCAGGTCAAGGTTTGCTACATTTTGTATGTGTGTGATGCTTCTCCCAAGGTATA	9974						
Qy	1642	TTAACTATATAGAGAGATTCTGACAAAACAGAAATGATAAAGCTGCGACCGTGGCACAG	1701						
Db	9975	TTAACTATATAGAGAGATTCTGACAAAACAGAAATGATAAAGCTGCGACCGTGGCACAG	10034						
Qy	1702	CTCATAGTTCTAGCTGCTTGGGAGTTTGAGGAGGAGGATGGCTTGAACACAGGTGTTC	1761						
Db	10035	CTCATAGTTCTAGCTGCTTGGGAGTTTGAGGAGGAGGATGGCTTGAACACAGGTGTTC	10094						
Qy	1762	AGCCAGCCTGGGCAACATAACAAGATCTCTCTCAAAAAAATAAAAAAAGAA	1821						
Db	10095	AGCCAGCCTGGGCAACATAACAAGATCTCTCTCAAAAAAATAAAAAAAGAA	10154						
Qy	1822	AGAGAGAGGCGCGGCTGGTGGCTCAGCGCTCTAATCCAGCACCTTTGGAGCGCGAGC	1881						
Db	10155	AGAGAGAGGCGCGGCTGGTGGCTCAGCGCTCTAATCCAGCACCTTTGGAGCGCGAGC	10214						
Qy	1882	CGGCGGATCACCTGTGTGTCAGGAGTTTGAGACACAGCTGCGCAACATGCAAAACCCCG	1941						
Db	10215	CGGCGGATCACCTGTGTGTCAGGAGTTTGAGACACAGCTGCGCAACATGCAAAACCCCG	10274						
Qy	1942	TCGTACTCAAAATGCAAAAATTTAGCAGCGCTGTAGCAGGACCTGTAAATCCAGCTA	2001						
Db	10275	TCGTACTCAAAATGCAAAAATTTAGCAGCGCTGTAGCAGGACCTGTAAATCCAGCTA	10334						
Qy	2002	CTTGGAGGCTGAGGAGGAGAAATCGCTTGAACCCAGGAGGTGGAGTTGCAGTAAGCTG	2061						
Db	10335	CTTGGAGGCTGAGGAGGAGAAATCGCTTGAACCCAGGAGGTGGAGTTGCAGTAAGCTG	10394						
Qy	2062	AGATCGTGGCGTTGCTACTCCAGCTGGGCGACAGAGACAGTCTGTCTCAGAAAAA	2121						
Db	10395	AGATCGTGGCGTTGCTACTCCAGCTGGGCGACAGAGACAGTCTGTCTCAGAAAAA	10454						
Qy	2122	AAAAAAAGAGAGAGAGAGAGAAAGACAATATTTGGGAGAGAAGGATGGGAAGCA	2181						
Db	10455	AAAAAAAGAGAGAGAGAGAGAAAGACAATATTTGGGAGAGAAGGATGGGAAGCA	10514						
Qy	2182	TTCAAGGAAATTTGCTTTTATCCAAATGTAAGGAGCCAAATAGGGATCCCTATTT	2241						
Db	10515	TTCAAGGAAATTTGCTTTTATCCAAATGTAAGGAGCCAAATAGGGATCCCTATTT	10574						
Qy	2242	GTCTCTTTTGGTGTCTATTTGTCCTTAAACACTGTCTTTGACAGTGAAGAAAAATATT	2301						
Db	10575	GTCTCTTTTGGTGTCTATTTGTCCTTAAACACTGTCTTTGACAGTGAAGAAAAATATT	10634						
Qy	2302	AATAACCATATCCTGTGCGGTTTATACCTAGCAACCCCTTGCAATGAAGATGACAGATC	2361						
Db	10635	AATAACCATATCCTGTGCGGTTTATACCTAGCAACCCCTTGCAATGAAGATGACAGATC	10694						
Qy	2362	CACAGGAAAACTTGAATGCACAACTGTCTTATTTTAACTCTTATTGTACATAAGTTCTAA	2421						
Db	10695	CACAGGAAAACTTGAATGCACAACTGTCTTATTTTAACTCTTATTGTACATAAGTTCTAA	10754						
Qy	2422	AAGAGTTAAAAATTTGTTACTTTCATGTATTCATTTATATTTTATTTTTCGCTCTAAT	2481						
Db	10755	AAGAGTTAAAAATTTGTTACTTTCATGTATTCATTTATATTTTATTTTTCGCTCTAAT	10814						
Qy	2482	GATTTTTTATACATGATTTTCTCTGATATATTGAAATGGAGTCTCAAGCTTCAT	2541						
Db	10815	GATTTTTTATACATGATTTTCTCTGATATATTGAAATGGAGTCTCAAGCTTCAT	10874						
Qy	2542	AAATTTATAACTTTAGAAATGATTCTAATAACACGATCTCTAATTTGTAACATTCACATA	2601						
Db	10875	AAATTTATAACTTTAGAAATGATTCTAATAACACGATCTCTAATTTGTAACATTCACATA	10934						

QY	2602	TGTCGTACGAAGCCATTTCTCTTGATTTTTAGTAAACTTTTATGACAGCAAAATTTGCTT	2661
Db	10935	TGTCGTACGAAGCCATTTCTCTTGATTTTTAGTAAACTTTTATGACAGCAAAATTTGCTT	10994
QY	2662	CTGGCTCACATTTCAATCAGTAAATAAATGATAAATAATTTTGGGAAGCTGTGAAGATAAA	2721
Db	10995	CTGGCTCACATTTCAATCAGTAAATAAATGATAAATAATTTTGGGAAGCTGTGAAGATAAA	11054
QY	2722	ATACCAATAAATAATAATAAAGTGATTTATATGAAGTTAAAAATAAAAAATCAGTATGA	2781
Db	11055	ATACCAATAAATAATAATAAAGTGATTTATATGAAGTTAAAAATAAAAAATCAGTATGA	11114
QY	2782	TGGAATAAAACTTGA	2795
Db	11115	TGGAATAAAACTTGA	11128
RESULT 3			
AC092184			
LOCUS			
DEFINITION	Homo sapiens 12 BAC RP11-438L7 (Roswell Park Cancer Institute Human BAC library) complete sequence.	71132 bp DNA linear	PRI 12-JUN-2002
ACCESSION	AC092184	AC013443	
VERSION	AC092184.7	GI:21206067	
KEYWORDS	HTG.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 71132)		
	Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayelle,M., Banks,T., Barbarta,J., Benton,J., Blinage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carton,F.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Emerling,S., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C., Hollins,B., Homsfi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Ioshikhes,I., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jollivet,S., Joudan,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Liu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Marondel,I., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mel,G., Merscher,S., Metabatt,M., Miller,A., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,M., Morris,S., Moser,M., Neal,D., Nelson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shim,C., Shoohtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,K., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wlecczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Kucherlapati,R., Weinstock,G. and Gibbs,R.		

STS /function="clone overlap"  
187..408  
/standard\_name="57233"  
STS 439..560  
/standard\_name="92005"  
774..881  
/rpt\_family="MIR"  
repeat\_region  
903..1190  
/rpt\_family="AluSq"  
repeat\_region  
1191..1213  
/rpt\_family="AT\_rich"  
STS 1591..1807  
/standard\_name="6612"  
STS 1744..1819  
/standard\_name="8198"  
repeat\_region  
1966..2264  
/rpt\_family="AluSq"  
repeat\_region  
3296..3328  
/rpt\_family=" (TTTC)n"  
repeat\_region  
4097..4249  
/rpt\_family="AluSq"  
repeat\_region  
4360..4654  
/rpt\_family="AluY"  
repeat\_region  
complement(5140..5262)  
/rpt\_family="FLAM\_C"  
5275..5304  
/rpt\_family="AT\_rich"  
repeat\_region  
5305..5563  
/rpt\_family="AluJo"  
repeat\_region  
5564..5599  
/rpt\_family="AT\_rich"  
repeat\_region  
5644..5733  
/rpt\_family="MSTD"  
repeat\_region  
5734..6012  
/rpt\_family="AluSq"  
repeat\_region  
6020..6336  
/rpt\_family="AluX"  
repeat\_region  
6337..6629  
/rpt\_family="MSTD"  
repeat\_region  
6630..6650  
/rpt\_family="(CAA)n"  
repeat\_region  
complement(6711..6817)  
/rpt\_family="MIR"  
repeat\_region  
complement(7033..7327)  
/rpt\_family="AluX"  
repeat\_region  
7607..7745  
/rpt\_family="FLAM\_A"  
repeat\_region  
7980..8047  
/rpt\_family="MIR"  
repeat\_region  
complement(8050..8549)  
/rpt\_family="LTR47A"  
repeat\_region  
8567..8698  
/rpt\_family="MIR"  
repeat\_region  
complement(8815..9101)  
/rpt\_family="AluSg"  
repeat\_region  
complement(11227..12175)  
/rpt\_family="LTR5"  
STS 12625..12900  
/standard\_name="87432"  
complement(13547..13664)  
/rpt\_family="L1MB5"  
repeat\_region  
complement(13665..13979)  
/rpt\_family="AluY"  
repeat\_region  
complement(13980..14193)  
/rpt\_family="L1MB5"  
repeat\_region  
14622..14924  
/rpt\_family="AluY"  
repeat\_region  
complement(15813..17136)  
/rpt\_family="L1P8"  
repeat\_region  
complement(17272..17395)  
/rpt\_family="FLAM\_C"  
repeat\_region  
17426..17575  
/rpt\_family="MIR"

repeat\_region 19021..19314  
Query Match 77.1%; Score 2174; DB 9; Length 71132;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 622 GCCCCTGTATGAGGTTGATGACTTTACGAGCGCATTTTCGTACTTTGGGACTTTGATAGCA 681  
DB 43498 GCCCCTGTATGAGGTTGATGACTTTACGAGCGCATTTTCGTACTTTGGGACTTTGATAGCA 43557  
QY 682 ACTTCCAGGAATGTCACACAGGATGAATATCTCTCTGAAGACAGTGGATAAAAACAG 741  
DB 43558 ACTTCCAGGAATGTCACACAGGATGAATATCTCTCTGAAGACAGTGGATAAAAACAG 43617  
QY 742 TCCTTCAAGTCTTCTCTGTTTTTATTCTTCAACCTCTCACCTTTCTTCTAGAGTTTACAGAAA 801  
DB 43618 TCCTTCAAGTCTTCTCTGTTTTTATTCTTCAACCTCTCACCTTTCTTCTAGAGTTTACAGAAA 43677  
QY 802 AATATTTATATACGACTCTTTTAAAGATCTATCTTGAAGATAGAGAGGACACAGG 861  
DB 43678 AATATTTATATACGACTCTTTTAAAGATCTATCTTGAAGATAGAGAGGACACAGG 43737  
QY 862 TCTGGCCAGGACGCTGCTGCAATTTGGTGCAGTTTGAATGCAACATTTGTCCTTCTACTGGG 921  
DB 43738 TCTGGCCAGGACGCTGCTGCAATTTGGTGCAGTTTGAATGCAACATTTGTCCTTCTACTGGG 43797  
QY 922 AATAACAGAACTGCAGGACCTGGGAGCATCCTAAAGTGTCAACGTTTTTCTTATGACTTTT 981  
DB 43798 AATAACAGAACTGCAGGACCTGGGAGCATCCTAAAGTGTCAACGTTTTTCTTATGACTTTT 43857  
QY 982 AGTATGATGAGAGCAGAGGATAGATCCTAAAAAGCATGGTGAGAGGATCAAAATGTTTTT 1041  
DB 43858 AGTATGATGAGAGCAGAGGATAGATCCTAAAAAGCATGGTGAGAGGATCAAAATGTTTTT 43917  
QY 1042 ATATCAACATCCTTTATTTATTTGATTTCATTTAGCTTAACAGTGTGTAGTATGATAGATT 1101  
DB 43918 ATATCAACATCCTTTATTTATTTGATTTCATTTAGCTTAACAGTGTGTAGTATGATAGATT 43977  
QY 1102 TTCTATTCTTTTCCCTTGGAGTTTACTTTTCAAGTAACACAAACTCTTCCATCAGGCCATG 1161  
DB 43978 TTCTATTCTTTTCCCTTGGAGTTTACTTTTCAAGTAACACAAACTCTTCCATCAGGCCATG 44037  
QY 1162 ATCTATAGGACCTCCTAATGAGATATCTGGGTGATTGTGACCCCAACCACTCTCTCAA 1221  
DB 44038 ATCTATAGGACCTCCTAATGAGATATCTGGGTGATTGTGACCCCAACCACTCTCTCAA 44097  
QY 1222 AGCATTAATATCCAATCATCGCTGTATGTTTTAATCAGCAGAGCATGTTTTTATGTTT 1281  
DB 44098 AGCATTAATATCCAATCATCGCTGTATGTTTTAATCAGCAGAGCATGTTTTTATGTTT 44157  
QY 1282 GTACAAAAGAGATTGTTATGGGTGGGATAGAGGTATAGACCATGATGGTCACTTCA 1341  
DB 44158 GTACAAAAGAGATTGTTATGGGTGGGATAGAGGTATAGACCATGATGGTCACTTCA 44217  
QY 1342 AGCTACTTTTATAAAGGATCTTAAATGGGAGGAGGACTGTGAACAGACACCCCTAATA 1401  
DB 44218 AGCTACTTTTATAAAGGATCTTAAATGGGAGGAGGACTGTGAACAGACACCCCTAATA 44277  
QY 1402 ATGGGTTGATGCTCTGAAGTAGCAATCTTCTGGAAAGCGCAACTCTTTTAAAGGAAGTCCC 1461  
DB 44278 ATGGGTTGATGCTCTGAAGTAGCAATCTTCTGGAAAGCGCAACTCTTTTAAAGGAAGTCCC 44337  
QY 1462 TAATTTAGAAAACACCCCAAACTTCACATATCATAATTAGCAAACTTTGGAAGGAAGTT 1521  
DB 44338 TAATTTAGAAAACACCCCAAACTTCACATATCATAATTAGCAAACTTTGGAAGGAAGTT 44397  
QY 1522 GCTTGAATGTTGGGAGAGGAAAATCTATTGGCTCTCGTGGGTCTCTTCAATCTCAGAAAT 1581  
DB 44398 GCTTGAATGTTGGGAGAGGAAAATCTATTGGCTCTCGTGGGTCTCTTCAATCTCAGAAAT 44457  
QY 1582 GCCAATCAGGTCAAGGTTTGGTACATTTTGTATGTTGTATGCTTCTCCCAAGGTATA 1641  
DB 44458 GCCAATCAGGTCAAGGTTTGGTACATTTTGTATGTTGTATGCTTCTCCCAAGGTATA 44517

Qy	1642	TTAACTATATAGAGAGTTGTGACAAACACAGAAATGATAAGCTGCGAACCCGTGGCACAGC	1701
Db	44518	TTAACTATATAGAGAGTTGTGACAAACACAGAAATGATAAGCTGCGAACCCGTGGCACAGC	44577
Qy	1702	CTCATAGTTCTAGCTGCTTGGAGGTTGAGGAGGAGATGGCTTGAACACAGGTGTTC	1761
Db	44578	CTCATAGTTCTAGCTGCTTGGAGGTTGAGGAGGAGATGGCTTGAACACAGGTGTTC	44637
Qy	1762	AGCCAGCCTGGCAACATACAGATCCCTCTCTCAAAAAAAGAAAAAGAA	1821
Db	44638	AGCCAGCCTGGCAACATACAGATCCCTCTCTCAAAAAAAGAAAAAGAA	44697
Qy	1822	AGAGAGAGGCGGCGGTGGTGTGCTACGCCGTGTAATCCAGCACTTTGGAGGCGCAGC	1881
Db	44698	AGAGAGAGGCGGCGGTGGTGTGCTACGCCGTGTAATCCAGCACTTTGGAGGCGCAGC	44757
Qy	1882	CGGGGGATCACCCTGTGTGAGGAGTTTGAGACCAAGCTGGCCCAACATGGCAAAACCCCG	1941
Db	44758	CGGGGGATCACCCTGTGTGAGGAGTTTGAGACCAAGCTGGCCCAACATGGCAAAACCCCG	44817
Qy	1942	TCTGTACTCAAAATGCAAAAATAGCCAGGCGTGTAGCAGGCACCTGTATCCAGCTA	2001
Db	44818	TCTGTACTCAAAATGCAAAAATAGCCAGGCGTGTAGCAGGCACCTGTATCCAGCTA	44877
Qy	2002	CTTGGGAGGCTGAGGAGGAGAAATCGGTTGAACCCAGGAGTGGAGGTTGCAGTAAGCTG	2061
Db	44878	CTTGGGAGGCTGAGGAGGAGAAATCGGTTGAACCCAGGAGTGGAGGTTGCAGTAAGCTG	44937
Qy	2062	AGATCGTCCGCTTGCATCCAGCTGGCGACAGACAGCAAGCTCTGTCTCAGAAAAAA	2121
Db	44938	AGATCGTCCGCTTGCATCCAGCTGGCGACAGACAGCAAGCTCTGTCTCAGAAAAAA	44997
Qy	2122	AAAAAAAAGAGAGAGAGAGAGAAAGAGAAATATTTGGGAGAGAGGATGGGGAAGCA	2181
Db	44998	AAAAAAAAGAGAGAGAGAGAGAAAGAGAAATATTTGGGAGAGAGGATGGGGAAGCA	45057
Qy	2182	TTGCAAGAAATGTGCTTTATCCACAAATGTAAAGAGCCAAATAAGGATCCCTATTT	2241
Db	45058	TTGCAAGAAATGTGCTTTATCCACAAATGTAAAGAGCCAAATAAGGATCCCTATTT	45117
Qy	2242	GTCTCTTTTGTGTCTATTTGTCCCTAACACTGCTTTTGCACAGTGAGAAAAATATTCAG	2301
Db	45118	GTCTCTTTTGTGTCTATTTGTCCCTAACACTGCTTTTGCACAGTGAGAAAAATATTCAG	45177
Qy	2302	AATAACCATATCCCTGTGCCGTTATTACCTAGCAACCCCTTGCAATGAAGATGAGCAGATC	2361
Db	45178	AATAACCATATCCCTGTGCCGTTATTACCTAGCAACCCCTTGCAATGAAGATGAGCAGATC	45237
Qy	2362	CACAGGAAACTTGAATGCACACTGCTTTATTTTATCTATTGTACATAAGTTTGTAA	2421
Db	45238	CACAGGAAACTTGAATGCACACTGCTTTATTTTATCTATTGTACATAAGTTTGTAA	45297
Qy	2422	AAGAGTTAAAAATTTTACTTTCATGTATTCATTTATATTTTATATTTTGGTCTCAAT	2481
Db	45298	AAGAGTTAAAAATTTTACTTTCATGTATTCATTTATATTTTATATTTTGGTCTCAAT	45357
Qy	2482	GATTTTTTATTAACATGATTTTCCCTTTTCTGATATATTGAATGGAGTCTCAAAAGCTTCAT	2541
Db	45358	GATTTTTTATTAACATGATTTTCCCTTTTCTGATATATTGAATGGAGTCTCAAAAGCTTCAT	45417
Qy	2542	AAATTTTAACTTTAGAAATGATTTCTAATAACACGATGATGTAATGTAACATTTGCAGTAA	2601
Db	45418	AAATTTTAACTTTAGAAATGATTTCTAATAACACGATGATGTAATGTAACATTTGCAGTAA	45477
Qy	2602	TGGTGTACGAAGCATTTCTCTGTATTTTGTATTTTGTAACTTTTATGACAGCAAAATTCCTT	2661
Db	45478	TGGTGTACGAAGCATTTCTCTGTATTTTGTATTTTGTAACTTTTATGACAGCAAAATTCCTT	45537
Qy	2662	CTGGCTCACCTTTCAATCAGTTAAATAAATGATAAATTTTGAAGCTCTGAAGATAAA	2721
Db	45538	CTGGCTCACCTTTCAATCAGTTAAATAAATGATAAATTTTGAAGCTCTGAAGATAAA	45597
Qy	2722	ATACCAATAAATAATAATAAAGCTGATTTATATGAAGTTTAAATAAAAAAATCAGTATGA	2781
Db	45598	ATACCAATAAATAATAATAAAGCTGATTTATATGAAGTTTAAATAAAAAAATCAGTATGA	45657
Qy	2782	TGGAATAAACTTGA	2795
Db	45658	TGGAATAAACTTGA	45671
RESULT 4			
LOCUS	BC006296	1837 bp	mRNA
DEFINITION	Homo sapiens, activation-induced cytidine deaminase, clone		
ACCESSION	MC:12911	IMAGE:4054915	mrna, complete cds.
VERSION	BC006296.1	GI:13623400	
KEYWORDS	MGC.		
SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 1837)		
JOURNAL	Strausberg, R.		
REMARK	Direct Submission		
COMMENT	Submitted (09-APR-2001) National Institutes of Health, Mammalian		
	Gene Collection (MGC), Cancer Genomics Office, National Cancer		
	Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,		
	USA		
	NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a>		
	Contact: MGC help desk		
	Email: <a href="mailto:cgabs-remail.nih.gov">cgabs-remail.nih.gov</a>		
	Tissue Procurement: Louis Staudt		
	cDNA Library Preparation: Rubin Laboratory		
	DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)		
	Sequencing Center (NISC),		
	Gaithersburg, Maryland:		
	Web site: <a href="http://www.nisc.nih.gov/">http://www.nisc.nih.gov/</a>		
	Contact: <a href="mailto:nisc.mgc@hgrl.nih.gov">nisc.mgc@hgrl.nih.gov</a>		
	Shenchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,		
	Benjamin, B., Blakesley, R.W., Bouffard, C.G., Brinkley, C., Brooks, S.,		
	Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Laquaspi, R.,		
	Lim, M., Maduro, O.L., Mastello, C., Mastrian, S.D., McCloskey, J.C.,		
	McDowell, J., Pearson, R., Snyder, B., Stantripop, S., Thomas, P.J.,		
	Tiongson, E.E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A.,		
	Zhang, L.-H. and Green, E.D.		
	Clone distribution: MGC clone distribution information can be found		
	through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a>		
	Series: IRAL Plate: 17 Row: a Column: 1		
	This clone was selected for full length sequencing because it		
	passed the following selection criteria: matched mRNA q1: 9984409.		
FEATURES	Location/Qualifiers		
source	1..1837		
	/organism: "Homo sapiens"		
	/db_xref: "LocusID:57379"		
	/db_xref: "taxon:9606"		
	/clone: "MGC:12911 IMAGE:4054915"		
	/tissue_type: "Primary B-Cells from Tonsils"		
	/clone_lib: "NIH_MGC_48"		
	/lab_host: "DH10B-R"		
	/note: "Vector: pOT87"		
	86...682		
	/codon_start: 1		
	/product: "activation-induced cytidine deaminase"		
	/protein_id: "AAH06296.1"		
	/db_xref: "GI:13623401"		
CDS			
	/translation: "MDSLIMNRKKFLYQFNVRWAKRRRETYICVYVKKRDSATSFSL		
	DFGYLRNKKVCEVLLFLRYISDWLDPCRCYRVTFWTSWSPCYDCARIVADPLKCNP		
	NLSLRFTFARLYFCEDRKAPEGLRLHRAGVQIAIMTFKDYTCWNTVENHRTFFK		
	ANEGLHNSVRLSRQLRLILPLVEVDDLRDAFRTLGL"		
BASE COUNT	530 a 387 c 421 g 499 t		
ORIGIN			

Query Match	64.4%	Score 1815;	DB 9;	Length 1837;
Best Local Similarity	100.0%;	Prod. No. 0;		
Matches 1815;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	4	GAACCATCATTAATTTGAAGTGGAGATTTTCTGGCCTGAGACTTTCGACGGAGGCAAGAAGA	63	
DB	10	GAACCATCATTAATTTGAAGTGGAGATTTTCTGGCCTGAGACTTTCGACGGAGGCAAGAAGA	69	
QY	64	CACCTCTGGACACCACTATGAGACGCCCTCTGTATGAACCGGAGGAAGTTTCTTTACCAATT	123	
DB	70	CACCTCTGGACACCACTATGAGACGCCCTCTGTATGAACCGGAGGAAGTTTCTTTACCAATT	129	
QY	124	CAAAAATGTCGCTGGCTAAGGTGGCTGAGACCTACCTGCTGCTACCTAGTGTGAAGAG	183	
DB	130	CAAAAATGTCGCTGGCTAAGGTGGCTGAGACCTACCTGCTGCTACCTAGTGTGAAGAG	189	
QY	184	GCCTGACAGTGTACATPCTTTTTCACCTGGACTTTTGGTTATCTTCGCAATTAAGAACGGCTG	243	
DB	190	GCCTGACAGTGTACATPCTTTTTCACCTGGACTTTTGGTTATCTTCGCAATTAAGAACGGCTG	249	
QY	244	CCAGCTGGAATTGCTCTTCTCGCTACATCTCGGACTGGACCTAGACCTGGCCGCTG	303	
DB	250	CCAGCTGGAATTGCTCTTCTCGCTACATCTCGGACTGGACCTAGACCTGGCCGCTG	309	
QY	304	CTACCGCGTCACCTGGTTCACCTCCTGGAGCCCTGCTACGACTGTGCCCGACATGTGGC	363	
DB	310	CTACCGCGTCACCTGGTTCACCTCCTGGAGCCCTGCTACGACTGTGCCCGACATGTGGC	369	
QY	364	CGACTTCTCGGAGGGAACCCAACTCAGTCTGAGGATCTTCACCGCGGCGCTCTACTT	423	
DB	370	CGACTTCTCGGAGGGAACCCAACTCAGTCTGAGGATCTTCACCGCGGCGCTCTACTT	429	
QY	424	CTGTGAGGACCGCAAGGCTGAGCCGAGGGCTGCGCGGCTGCACCGCGCGGGTGCA	483	
DB	430	CTGTGAGGACCGCAAGGCTGAGCCGAGGGCTGCGCGGCTGCACCGCGCGGGTGCA	489	
QY	484	AATAGCATCATGACCTTCAAAGATTAATTTTACTGCTGGAACTACTTTGTAGAAAAACCA	543	
DB	490	AATAGCATCATGACCTTCAAAGATTAATTTTACTGCTGGAACTACTTTGTAGAAAAACCA	549	
QY	544	TGAAGAACCTTTCAAAGCCTGGAAGGGCTGCATGAAATTCAGTTCGCTCTCCAGACA	603	
DB	550	TGAAGAACCTTTCAAAGCCTGGAAGGGCTGCATGAAATTCAGTTCGCTCTCCAGACA	609	
QY	604	GCCTTGGCGCATCCTTTTGGCCCTGTATGAGGTGTGACTTACGAGACGCAATTCGTAC	663	
DB	610	GCCTTGGCGCATCCTTTTGGCCCTGTATGAGGTGTGACTTACGAGACGCAATTCGTAC	669	
QY	664	TTTGGGACTTTGATAGCAACTTCCAGAAATGTCACACAGCATGAAATATCTGCTCAAG	723	
DB	670	TTTGGGACTTTGATAGCAACTTCCAGAAATGTCACACAGCATGAAATATCTGCTCAAG	729	
QY	724	ACAGTGGATAAAAAACAGTCTCTCAAGTCTCTCTGTTTATTTCTTCAACTCTCACTTT	783	
DB	730	ACAGTGGATAAAAAACAGTCTCTCAAGTCTCTCTGTTTATTTCTTCAACTCTCACTTT	789	
QY	784	CTTAGAGTTTACAGAAAAAATATTTATAGGACTCTTTAAAAAGATCTATGCTTGAAGA	843	
DB	790	CTTAGAGTTTACAGAAAAAATATTTATAGGACTCTTTAAAAAGATCTATGCTTGAAGA	849	
QY	844	ATAGAGAAGGAACACAGGTCTGGCCAGGGACGTGCTCAATTTGGTGCAGTTTTGAATGCA	903	
DB	850	ATAGAGAAGGAACACAGGTCTGGCCAGGGACGTGCTCAATTTGGTGCAGTTTTGAATGCA	909	
QY	904	ACATTTGCCCTTACTGGGAATAACAGAACTGCAGGACCTGGGACATCTCTAAAAGTGTCAA	963	
DB	910	ACATTTGCCCTTACTGGGAATAACAGAACTGCAGGACCTGGGACATCTCTAAAAGTGTCAA	969	
QY	964	CGTTTTTCTATGACTTTTAGGTAGGTAGTACAGACAGAGGTAGATCTCTAAAAGCATGGTG	1023	
DB	970	CGTTTTTCTATGACTTTTAGGTAGGTAGTACAGACAGAGGTAGATCTCTAAAAGCATGGTG	1029	

REFERENCE 2 (bases 1 to 596)  
AUTHORS Martin.A. and Scharff,M.D.  
TITLE Direct Submission  
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA  
FEATURES  
source  
1..596  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="2"  
/cell\_type="hybridoma P1-5"  
/transgenic  
/transgenic  
1..596  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
1..>596  
/gene="AID"  
1..>596  
/gene="AID"  
/codon\_start=1  
/product="AID"  
/protein\_id="AAM95416.1"  
/db\_xref="GI:22297246"  
/translation="MDSLMNRKFLYOFKNVRWAKGRRETYLCYVVKRRDSATSFSL  
DFGYLRNKGCHVELLFURYISDWLDPGRCYRVTWFTSWSPCYDCARHVDLFRGNP  
NLSLRIFTARLYFCEDRKAEPGLRRLHRAVQIAIMTFKDYFCWNTFVENHIERTEK  
AWEGLHENSVRSLRLLPLLYEVDLDRDAFRTLGL"  
BASE COUNT 128 a 163 c 155 g 150 t  
ORIGIN  
Query Match 20.8%; Score 587; DB 12; Length 596;  
Best Local Similarity 100.0%; Pred. No. 1.2e-294;  
Matches 587; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 89 CTCCTGATGAACCGGAGGAGTTCTTTACCAATTCAAAATGTCCGCTGGCTAAGGT 148  
Db 10 CTCCTGATGAACCGGAGGAGTTCTTTACCAATTCAAAATGTCCGCTGGCTAAGGT 69  
QY 149 CGCGCTGAGACCTACCTGTGTACGTAGTGAAGAGCGGTGACAGTGTACATCTCTTTCA 208  
Db 70 CGCGCTGAGACCTACCTGTGTACGTAGTGAAGAGCGGTGACAGTGTACATCTCTTTCA 129  
QY 209 CTGGACTTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGAATTCCTTCCTCGC 268  
Db 130 CTGAGCTTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGAATTCCTTCCTCGC 189  
QY 269 TACATCTCGGACTGGGACCTAGACCCGTGCGCGCTGCTACCGCGTCAACCTGTTCACTCC 328  
Db 190 TACATCTCGGACTGGGACCTAGACCCGTGCGCGCTGCTACCGCGTCAACCTGTTCACTCC 249  
QY 329 TGGAGCCCTGCTACGACTGTGCCGACATGTGGCCGACTTTCTGCGAGGGAACCCCAAC 388  
Db 250 TGGAGCCCTGCTACGACTGTGCCGACATGTGGCCGACTTTCTGCGAGGGAACCCCAAC 309  
QY 389 CTCAGTCTGAGGATCTTCACGCGCGCTCTACTTCTGTGAGGACCGCAAGCTGAGGCC 448  
Db 310 CTCAGTCTGAGGATCTTCACGCGCGCTCTACTTCTGTGAGGACCGCAAGCTGAGGCC 369  
QY 449 GAGGGCTGCGCGCGCTGACCGCGCGGGGTGCAAAATAGCCATCATGACCTTCAAGAT 508  
Db 370 GAGGGCTGCGCGCGCTGACCGCGCGGGGTGCAAAATAGCCATCATGACCTTCAAGAT 429  
QY 509 TATTTTTACTGCTGGAATACTTTTGTAGAAACCATGAAGAACATTTCAAAGCCTGGAA 568  
Db 430 TATTTTTACTGCTGGAATACTTTTGTAGAAACCATGAAGAACATTTCAAAGCCTGGAA 489  
QY 569 GGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTCGGCGCATCCTTTGGCCCTG 628  
Db 490 GGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTCGGCGCATCCTTTGGCCCTG 549  
QY 629 TATGAGTTGATGACTTACGAGACGATTTTCGTACTTTGGGACTTTG 675  
Db 550 TATGAGTTGATGACTTACGAGACGATTTTCGTACTTTGGGACTTTG 596

RESULT 6  
AF529823  
LOCUS Homo sapiens clone Ramos 9 AID (AID) mRNA linear PRI 19-AUG-2002  
DEFINITION Homo sapiens clone Ramos 9 AID (AID) mRNA, partial cds.  
ACCESSION AF529823  
VERSION AF529823.1 GI:22297233  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 596)  
AUTHORS Martin.A. and Scharff,M.D.  
TITLE Somatic hypermutation of the AID transgene in B cells and non-B cells  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 596)  
AUTHORS Martin.A. and Scharff,M.D.  
TITLE Direct Submission  
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA  
FEATURES  
source  
1..596  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Ramos 9"  
1..>596  
/gene="AID"  
1..>596  
/gene="AID"  
/note="Integrated into Burkitt's lymphoma cell line Ramos"  
/codon\_start=1  
/product="AID"  
/protein\_id="AAM95410.1"  
/db\_xref="GI:22297234"  
BASE COUNT 128 a 163 c 155 g 150 t  
ORIGIN  
Query Match 20.7%; Score 584; DB 9; Length 596;  
Best Local Similarity 100.0%; Pred. No. 4.5e-293;  
Matches 584; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 92 TTGATGAACCGGAGGAGTTCTTTTACCAATTCAAAATGTCCGCTGGCTAAGGTGCG 151  
Db 13 TTGATGAACCGGAGGAGTTCTTTTACCAATTCAAAATGTCCGCTGGCTAAGGTGCG 72  
QY 152 CGTGAGACCTACCTGCTAGCTAGTGAAGAGCGGTGACAGTGTACATCTTTTCACTG 211  
Db 73 CGTGAGACCTACCTGCTAGCTAGTGAAGAGCGGTGACAGTGTACATCTTTTCACTG 132  
QY 212 GACTTTGTTTATCTTCGCAATAAGAACGGCTGCCACGTGGAATTCCTTCCTCGCTAC 271  
Db 133 GACTTTGTTTATCTTCGCAATAAGAACGGCTGCCACGTGGAATTCCTTCCTCGCTAC 192  
QY 272 ATCTCGGACTGGGACCTAGACCTGGCGCTGTACCTGCCGTGCTGTTGACCTTCGCG 331  
Db 193 ATCTCGGACTGGGACCTAGACCTGGCGCTGTACCTGCCGTGCTGTTGACCTTCCTCG 252  
QY 332 AGCCCTCTACGACTGTGCCGACATGTGCCGACTTTCTCGGAGGGAACCCCAACCTG 391  
Db 253 AGCCCTCTACGACTGTGCCGACATGTGCCGACTTTCTCGGAGGGAACCCCAACCTG 312  
QY 392 AGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGCAAGGTTCAGTCCGAG 451  
Db 313 AGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGCAAGGTTCAGTCCGAG 472  
QY 452 GGGCTGCGCGCGCTGACACCGCGCGGGGTGCAAAATAGCCATCATGACCTTCAAGATTTAT 511

```

Db      373 GGGCTGCGCGGCTGCACCGCGCGGGTGCAAAATAGCATCATCACCTTCAAAGATTAT 432
QY      512 TTTTACTGCTGGAATACCTTTTGTAGAAAACCATGAAGAACTTTCAAAGCCTGGGAAGG 571
Db      433 TTTTACTGCTGGAATACCTTTTGTAGAAAACCATGAAGAACTTTCAAAGCCTGGGAAGG 492
QY      572 CTGCATGAAAATTCAGTTTCGTCTCTCCAGACAGCTTCGGCGCATCCTTTTGCCTCTGTAT 631
Db      493 CTGCATGAAAATTCAGTTTCGTCTCTCCAGACAGCTTCGGCGCATCCTTTTGCCTCTGTAT 552
QY      632 GAGTTGTAGTACTTACGAGAGCATTTTCGTACTTTGGGACTTTG 675
Db      553 GAGTTGTAGTACTTACGAGAGCATTTTCGTACTTTGGGACTTTG 596

RESULT 7
AF529846
LOCUS   AF529846
DEFINITION Cricetulus griseus clone 597 bp mRNA linear SYN 19-AUG-2002
complete cds.
ACCESSION AF529846
VERSION   AF529846.1 GI:22297279
SOURCE    Chinese hamster.
ORGANISM  Cricetulus griseus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
           Cricetulus;
REFERENCE 1 (bases 1 to 597)
AUTHORS   Martin,A. and Scharff,M.D.
TITLE     Somatic hypermutation of the AID transgene in B cells and non-B
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 597)
AUTHORS   Martin,A. and Scharff,M.D.
TITLE     Direct Submission
JOURNAL   Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
Source    1..597
           /organism="Cricetulus griseus"
           /db_xref="taxon:10029"
           /clone="6"
           /cell_line="CHO"
           /transgenic
source    1..597
           /organism="Homo sapiens"
           /db_xref="taxon:9606"
gene      1..597
           /gene="AID"
CDS       1..597
           /gene="AID"
           /codon_start=1
           /product="AID"
           /protein_id="AA095433.1"
           /db_xref="GI:22297280"
           /translation="MDSLIMNRKFLYQFNVRWAKRRRTYLCVWVKRRDSATFSFL
           DFGYLNRKNGCHVELLELYRISDWLDPRGCVRYTWFTSWSPCYCARHVAFLRGNP
           NLSLRIFATRYFCEDRAEPEGLRLHRAGVQIAIMTFKDYFYCWNTFVENHRTFK
           AWGLJHENSRLVRSQRLRLILPLYEVDLDRAFRWGR"
BASE COUNT 128 a 165 c 157 g 147 t
ORIGIN

Query Match 20.3%; Score 572; DB 12; Length 597;
Best Local Similarity 100.0%; Pred. No. 8.2e-287;
Matches 572; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      94 GATGAACCGGAGGAAGTTTCTTTTACCAATTCAAAATGTCCTGGCTAAGGTCGGCG 153
Db      15 GATGAACCGGAGGAAGTTTCTTTTACCAATTCAAAATGTCCTGGCTAAGGTCGGCG 74
QY      154 TGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGCTACATCCTTTTCACTGGA 213
```

```

Db      75 TGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGCTACATCCTTTTCACTGGA 134
QY      214 CTTTGGTTATCTTCGCATAGAAGCGCTGCCAGTGAATTCCTTCCTCCGCTACAT 273
Db      135 CTTTGGTTATCTTCGCATAGAAGCGGTGCCAGTGAATTCCTTCCTCCGCTACAT 194
QY      274 CTCGAGCTGGGACCTAGACCCCTGGCCGCTGTACCGGCTCACCTGGTTTCACCTCCTCGGAG 333
Db      195 CTCGAGCTGGGACCTAGACCCCTGGCCGCTGTACCGGCTCACCTGGTTTCACCTCCTCGGAG 254
QY      334 CCCCTGCTACGACTGTGCCCGACATGTGGCGGACTTTCTGCGAGGGAACCCCAACCTCAG 393
Db      255 CCCCTGCTACGACTGTGCCCGACATGTGGCGGACTTTCTGCGAGGGAACCCCAACCTCAG 314
QY      394 TCTGAGGATCTTCACCGCGGCCCTCTACTTCTGTGAGGACCGCAAGGCTGAGCCCGAGGG 453
Db      315 TCTGAGGATCTTCACCGCGGCCCTCTACTTCTGTGAGGACCGCAAGGCTGAGCCCGAGGG 374
QY      454 GCTCGCGGCTGCACCGCGCGGGGTGCAATAGCCATCATGACCTTTCAAAGATTATTT 513
Db      375 GCTCGCGGCTGCACCGCGCGGGGTGCAATAGCCATCATGACCTTTCAAAGATTATTT 434
QY      514 TTACTGCTGGAATACTTTTGTAGAAAACCATGAAGAATCTTTCAAAGCCTGGGAAGGCT 573
Db      435 TTACTGCTGGAATACTTTTGTAGAAAACCATGAAGAATCTTTCAAAGCCTGGGAAGGCT 494
QY      574 GCATGAAAATTCAGTTCGTCTCTCCAGACAGCTTCGGCGCATCCTTTTGGCCCTGTATGA 633
Db      495 GCATGAAAATTCAGTTCGTCTCTCCAGACAGCTTCGGCGCATCCTTTTGGCCCTGTATGA 554
QY      634 GGTGTGATGACTTACGAGAGCATTTTCGTACTT 665
Db      555 GGTGTGATGACTTACGAGAGCATTTTCGTACTT 586

RESULT 8
AF529842
LOCUS   AF529842
DEFINITION Cricetulus griseus clone 597 bp mRNA linear SYN 19-AUG-2002
complete cds.
ACCESSION AF529842
VERSION   AF529842.1 GI:22297271
KEYWORDS
SOURCE    Chinese hamster.
ORGANISM  Cricetulus griseus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
           Cricetulus;
REFERENCE 1 (bases 1 to 597)
AUTHORS   Martin,A. and Scharff,M.D.
TITLE     Somatic hypermutation of the AID transgene in B cells and non-B
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 597)
AUTHORS   Martin,A. and Scharff,M.D.
TITLE     Direct Submission
JOURNAL   Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source    1..597
           /organism="Cricetulus griseus"
           /db_xref="taxon:10029"
           /clone="2"
           /cell_line="CHO"
           /transgenic
source    1..597
           /organism="Homo sapiens"
           /db_xref="taxon:9606"
gene      1..597
           /gene="AID"
CDS       1..597
           /gene="AID"
```

```
/codon_start=1
/product="AID"
/protein_id="AAM95429.1"
/db_xref="GI:22297272"
/translation="MDSLMMNRKFLYOFKNVRWAKGRRETYLCYVVKRRDSATSFSL
DFGLRNKNGCHVELLFURYISDWDLDPGRCYRVTWFTSWSPCYDCARHVAADFURGNP
NLSLRIFTFARLYFCEDRKAEPGLRLRHAGVQIAIMTFKDYFYCWNTFVENHIERTEK
AWEGLHNSVRLSRQLRLLPLVEVDLDRDAFTWGR"
BASE COUNT      127 a  164 c  157 g  149 t
ORIGIN
Query Match      20.2%; Score 569; DB 12; Length 597;
Best Local Similarity 100.0%; Pred. No. 3e-285;
Matches 569; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 80 ATGGACACCCCTTCATGAACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGGCTGG 139
Db 1 ATGGACACCCCTTCATGAACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGGCTGG 60
QY 140 GCTAAGGTCGGCGTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGTCTACA 199
Db 61 GCTAAGGTCGGCGTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGTCTACA 120
QY 200 TCCTTTTTCACGTGACCTTTGGTTATCTTCGCAATAAGAACGGCTGCCAGTGAATGTCTC 259
Db 121 TCCTTTTTCACGTGACCTTTGGTTATCTTCGCAATAAGAACGGCTGCCAGTGAATGTCTC 180
QY 260 TTCCCTCCGCTACATCTCGGACCTGCGACCTAGACCCCTGGCGCGTGTACCGGTACCTGG 319
Db 181 TTCCCTCCGCTACATCTCGGACCTGCGACCTAGACCCCTGGCGCGTGTACCGGTACCTGG 240
QY 320 TTCACCTCTCGAGCCCTGCTACGACTGTGCCGACATGTGGCGGACTTTCTTGGAGGG 379
Db 241 TTCACCTCTCGAGCCCTGCTACGACTGTGCCGACATGTGGCGGACTTTCTTGGAGGG 300
QY 380 RACCCCAACCTCAGTCTGAGNATCTTACCGGGCGCCCTCTACTTCTGTGAGGACCGCAAG 439
Db 301 RACCCCAACCTCAGTCTGAGNATCTTACCGGGCGCCCTCTACTTCTGTGAGGACCGCAAG 360
QY 440 GCTGAGCCGAGGGGCTGGCGGCTGCACCGCGCGGGTGCAATAGCCATCATGACC 499
Db 361 GCTGAGCCGAGGGGCTGGCGGCTGCACCGCGCGGGTGCAATAGCCATCATGACC 420
QY 500 TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAGAATTTTCAA 559
Db 421 TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAGAATTTTCAA 480
QY 560 GCCTGGGAAGGCTGCATGAATAATTCAGTTGCTCTCCAGACACCTTCGGGGCATCCTT 619
Db 481 GCCTGGGAAGGCTGCATGAATAATTCAGTTGCTCTCCAGACACCTTCGGGGCATCCTT 540
QY 620 TTGCCCCCTGTATGAGTTGATGACTTACG 648
Db 541 TTGCCCCCTGTATGAGTTGATGACTTACG 569
RESULT 9
AF529847
LOCUS      AF529847      597 bp      mRNA      linear      SVN 19-AUG-2002
DEFINITION Cricetulus griseus clone 7 transgenic Homo sapiens AID (AID) mRNA,
complete cds.
ACCESSION  AF529847
VERSION     AF529847.1
KEYWORDS   GI:22297281
SOURCE     Chinese hamster.
ORGANISM   Cricetulus griseus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
Cricetulus.
REFERENCE  1 (bases 1 to 597)
AUTHORS   Martin,A. and Scharff,M.D.
TITLE     Somatic hypermutation of the AID transgene in B cells and non-B
cells
```

```
Unpublished
REFERENCE    2 (bases 1 to 597)
AUTHORS     Martin,A. and Scharff,M.D.
TITLE       Direct Submission
JOURNAL     Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
Location/Qualifiers
1..597
/organism="Cricetulus griseus"
/db_xref="taxon:10029"
/clone="7"
/cell_line="CHO"
/transgenic
1..597
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..597
/gene="AID"
1..597
/gene="AID"
/protein_id="AAM95434.1"
/product="AID"
/codon_start=1
/db_xref="GI:22297282"
/translation="MDSLMMNRKFLYOFKNVRWAKGRRETYLCYVVKRRDSATSFSL
DFGLRNKNGCHVELLFURYISDWDLDPGRCYRVTWFTSWSPCYDCARHVAADFURGNP
NLSLRIFTFARLYFCEDRKAEPGLRLRHAGVQIAIMTFKDYFYCWNTFVENHIERTEK
AWEGLHNSVRLSRQLRLLPLVEVDLDRDAFTWGR"
BASE COUNT      129 a  164 c  156 g  148 t
ORIGIN
Query Match      19.4%; Score 548; DB 12; Length 597;
Best Local Similarity 100.0%; Pred. No. 2.7e-274;
Matches 548; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 80 ATGGACACCCCTTCGATGAACCGGAGGAAGTTTCTTTACCAATTCAAAAATGTCGGCTGG 139
Db 1 ATGGACACCCCTTCGATGAACCGGAGGAAGTTTCTTTACCAATTCAAAAATGTCGGCTGG 60
QY 140 GCTAAGGTCGGCGTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGTCTACA 199
Db 61 GCTAAGGTCGGCGTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGTCTACA 120
QY 200 TCCTTTTTCACGTGACCTTTGGTTATCTTCGCAATAAGAACGGCTGCCAGTGAATGTCTC 259
Db 121 TCCTTTTTCACGTGACCTTTGGTTATCTTCGCAATAAGAACGGCTGCCAGTGAATGTCTC 180
QY 260 TTCCCTCCGCTACATCTCGGACCTGCGACCTAGACCCCTGGCGCGTGTCTACCTACCTGG 319
Db 181 TTCCCTCCGCTACATCTCGGACCTGCGACCTAGACCCCTGGCGCGTGTCTACCTACCTGG 240
QY 320 TTCACCTCTCGAGCCCTGCTACGACTGTGCCGACATGTGGCGGACTTTCTTGGAGGG 379
Db 241 TTCACCTCTCGAGCCCTGCTACGACTGTGCCGACATGTGGCGGACTTTCTTGGAGGG 300
QY 380 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCCCTCTACTTCTGTGAGGACCGCAAG 439
Db 301 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCCCTCTACTTCTGTGAGGACCGCAAG 360
QY 440 GCTGAGCCGAGGGGCTGGCGGCTGCACCGCGCGGGTGCAATAGTATGACCTTCATGACC 499
Db 361 GCTGAGCCGAGGGGCTGGCGGCTGCACCGCGCGGGTGCAATAGTATGACCTTCATGACC 420
QY 500 TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAGAATTTTCAA 559
Db 421 TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAGAATTTTCAA 480
QY 560 GCCTGGGAAGGCTGCATGAATAATTCAGTTGCTCTCTCCAGACACCTTCGGGGCATCCTT 619
Db 481 GCCTGGGAAGGCTGCATGAATAATTCAGTTGCTCTCTCCAGACACCTTCGGGGCATCCTT 540
QY 620 TTGCCCCCT 627
|||||||
```



Db	541	TTGCCCT 548	
RESULT 10			
AF529828			
LOCUS	597 bp	mrna	linear
DEFINITION	Mus musculus clone 1 transgenic Homo sapiens AID (AID) mRNA,		SYN 19-AUG-2002
ACCESSION	AF529828		
VERSION	AF529828.1	GI:22297243	
KEYWORDS	house mouse,		
SOURCE	Mus musculus		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	1 (bases 1 to 597)		
AUTHORS	Martin,A. and Scharff,M.D.		
TITLE	Somatic hypermutation of the AID transgene in B and non-B cells		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 597)		
AUTHORS	Martin,A. and Scharff,M.D.		
TITLE	Direct Submission		
JOURNAL	Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA		
FEATURES	Location/Qualifiers		
source	1..597		
	/organism="Mus musculus"		
	/db_xref="taxon:10090"		
	/clone="1"		
	/cell_type="hybridoma PI-5"		
	/transgenic		
source	1..597		
	/gene="AID"		
	/codon_start=1		
	/product="AID"		
	/protein_id="AAM95415.1"		
gene	1..597		
	/db_xref="taxon:9606"		
	/gene="AID"		
CDS	1..597		
	/gene="AID"		
	/codon_start=1		
	/product="AID"		
	/protein_id="AAM95415.1"		
	/db_xref="GI:22297244"		
	/translation="MDSLMMNRKFLYQFNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNDCHVELLPFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVAADFRLGNPNLSLRIFTALYFCEDRKAEPEGLRLRHAGVOIAIMTFKDYFCYWNTEFVENHERTFKAWEGLHENSVELSRQLRILLPLLYEVDLDAFRTLGL"		
BASE COUNT	129 a	164 c	154 g
ORIGIN			
	19.4%; Score 546; DB 12; Length 597;		
Query Match	Best Local Similarity	99.8%;	Pred. No. 3e-273;
Matches	596; Conservative	0; Mismatches	1; Indels
	0; Gaps		0;
Qy	80	ATGGACAGCCTTCTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAATATGCCGCTGG	139
Db	1	ATGGACAGCCTTCTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAATATGCCGCTGG	60
Qy	140	GCTAAGGTCGCGCTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGCTACA	199
Db	61	GCTAAGGTCGCGCTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGCTACA	120
Qy	200	TCTTTTACTGGACTTTGGTTATCTTCGCAATGAAGACGCGTCCACGTTGGAATTGCTC	259
Db	121	TCTTTTACTGGACTTTGGTTATCTTCGCAATGAAGACGCGTCCACGTTGGAATTGCTC	180
Qy	260	TTCTCCCGTACATCTCGAGCTGGGACCTAGACCTGCGCGCTGCTACCGCGTACCTGG	319
Db	181	TTCTCCCGTACATCTCGAGCTGGGACCTAGACCTGCGCGCTGCTACCGCGTACCTGG	240
Qy	320	TTCACTCTCTGAGCCCTGCTACGACTGTGCCCGACATGTGCCGACTTCTTCTGCCAGGG	379
Db	241	TTCACTCTCTGAGCCCTGCTACGACTGTGCCCGACATGTGCCGACTTCTTCTGCCAGGG	300

Db 61 GCTAAGGTCGGCTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGCTACA 120  
 QY 200 TCTTTTCACTGACTTTGGTTATCTTCGCAATAGAACGGCTGCCACGTGAATTGCTC 259  
 Db 121 TCTTTTCACTGACTTTGGTTATCTTCGCAATAGAACGGCTGCCACGTGAATTGCTC 180  
 QY 260 TTCTCTCGCTACATCTCGGACTGGGACCTAGACCTGGCCGCTGTACCGCGTCACTGG 319  
 Db 181 TTCTCTCGCTACATCTCGGACTGGGACCTAGACCTGGCCGCTGTACCGCGTCACTGG 240  
 QY 320 TTCACTCTCTGAGCCCTGCTACGACTGTGCCGACATGTGGCGGACTTTCTGGAGGG 379  
 Db 241 TTCACTCTCTGAGCCCTGCTACGACTGTGCCGACATGTGGCGGACTTTCTGGAGGG 300  
 QY 380 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 439  
 Db 301 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360  
 QY 440 GCTGAGCCGAGGGCTGGCGGCTGACCGCGCGGGTGCACATAGCCATCATGACC 499  
 Db 361 GCTGAGCCGAGGGCTGGCGGCTGACCGCGCGGGTGCACATAGCCATCATGACC 420  
 QY 500 TTCAAGATATTTTACTGCTGGAATCTTTGTAGAAACCATCAAGAACTTTCRAA 559  
 Db 421 TTCAAGATATTTTACTGCTGGAATCTTTGTAGAAACCATCAAGAACTTTCRAA 480  
 QY 560 GCTGGGAAGGCTGCATGAAATTCAGTCTCTCCAGACAGCTTCCGCGCATCTT 619  
 Db 481 GCTGGGAAGGCTGCATGAAATTCAGTCTCTCCAGACAGCTTCCGCGCATCTT 540  
 QY 620 TTGCCCCCTATGAGGTTGATGACTTACGAGACGATTCGTACTTTGGGACTTTG 675  
 Db 541 TTGCCCCCTATGAGGTTGATGACTTACGAGACGATTCGTACTTTGGGACTTTG 596

RESULT 12  
 AF529816 Homo sapiens clone Ramos 2 AID (AID) mRNA linear PRI 19-AUG-2002  
 LOCUS AF529816  
 DEFINITION AF529816  
 ACCESSION AF529816  
 VERSION AF529816.1 GI:22297219  
 KEYWORDS human.  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 Martin, A. and Scharff, M.D.  
 Somatic hypermutation of the AID transgene in B cells and non-B  
 cells  
 Unpublished  
 2 (bases 1 to 596)  
 Martin, A. and Scharff, M.D.  
 Direct Submission  
 Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of  
 Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA  
 Location/Qualifiers  
 1. .596  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="Ramos 2"  
 1. .>596  
 /gene="Aid"  
 1. .>596  
 /note="Integrated into Burkitt's lymphoma cell line Ramos"  
 /codon\_start=1  
 /product="AID"  
 /protein\_id="RAM95403.1"  
 /db\_xref="GI:22297220"  
 /translation="MDSLMMNRKFLYQKNVRWAKGRRETYLCYVVKRRDSATSFSL  
 DFGYLRNKGCHVFLRYISDWLDPGRYRVTWFTSRSPCYDCARHVDALRGNP  
 NLSLRIFARLYFCEDRAKEPEGLRRLRHAGVQIAIMTFKDYFYCWNTFVENHETFK

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 FEATURES  
 source

AF529819 Homo sapiens clone Ramos 5 AID (AID) mRNA, partial cds.  
 AF529819  
 AF529819.1 GI:22297225  
 human.  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 596)  
 Martin, A. and Scharff, M.D.  
 Somatic hypermutation of the AID transgene in B cells and non-B  
 cells  
 Unpublished  
 2 (bases 1 to 596)  
 Martin, A. and Scharff, M.D.  
 Direct Submission  
 Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of  
 Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA  
 Location/Qualifiers  
 1. .596  
 /organism="Homo sapiens"

BASE COUNT 127 a 165 c 155 g 149 t  
 ORIGIN  
 Query Match 19.3%; Score 545; DB 9; Length 596;  
 Best Local Similarity 99.8%; Pred. No. 9.8e-273;  
 Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 80 ATGACAGCCCTCTGATGAACCGGAGGAAGTTTCTTTACCAATTCAAAAATGTCGGCTGG 139  
 Db 1 ATGACAGCCCTCTGATGAACCGGAGGAAGTTTCTTTACCAATTCAAAAATGTCGGCTGG 60  
 QY 140 GCTAAGGTCGGCGTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGCTACA 199  
 Db 61 GCTAAGGTCGGCGTGAGACCTACCTGTGCTAGCTAGTGAAGAGCGGTGACAGTGCTACA 120  
 QY 200 TCTTTTCACTGAGACTTTGGTTATCTTCGCAATAGAACGGCTGCCACGTGAATTGCTC 259  
 Db 121 TCTTTTCACTGAGACTTTGGTTATCTTCGCAATAGAACGGCTGCCACGTGAATTGCTC 180  
 QY 260 TTCTCTCGCTACATCTCGGACTGGGACCTAGACCTGGCCGCTGTACCGCGTCACTGG 319  
 Db 181 TTCTCTCGCTACATCTCGGACTGGGACCTAGACCTGGCCGCTGTACCGCGTCACTGG 240  
 QY 320 TTCACTCTCTGAGCCCTGCTACGACTGTGCCGACATGTGGCGGACTTTCTGGAGGG 379  
 Db 241 TTCACTCTCTGAGCCCTGCTACGACTGTGCCGACATGTGGCGGACTTTCTGGAGGG 300  
 QY 380 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 439  
 Db 301 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360  
 QY 440 GCTGAGCCGAGGGCTGGCGGCTGACCGCGCGGGTGCACATAGCCATCATGACC 499  
 Db 361 GCTGAGCCGAGGGCTGGCGGCTGACCGCGCGGGTGCACATAGCCATCATGACC 420  
 QY 500 TTCAAGATATTTTACTGCTGGAATCTTTGTAGAAACCATCAAGAACTTTCRAA 559  
 Db 421 TTCAAGATATTTTACTGCTGGAATCTTTGTAGAAACCATCAAGAACTTTCRAA 480  
 QY 560 GCTGGGAAGGCTGCATGAAATTCAGTCTCTCCAGACAGCTTCCGCGCATCTT 619  
 Db 481 GCTGGGAAGGCTGCATGAAATTCAGTCTCTCCAGACAGCTTCCGCGCATCTT 540  
 QY 620 TTGCCCCCTATGAGGTTGATGACTTACGAGACGATTCGTACTTTGGGACTTTG 675  
 Db 541 TTGCCCCCTATGAGGTTGATGACTTACGAGACGATTCGTACTTTGGGACTTTG 596

RESULT 13  
 AF529819  
 LOCUS AF529819  
 DEFINITION Homo sapiens clone Ramos 5 AID (AID) mRNA, partial cds.  
 ACCESSION AF529819  
 VERSION AF529819.1 GI:22297225  
 KEYWORDS human.  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 596)  
 Martin, A. and Scharff, M.D.  
 Somatic hypermutation of the AID transgene in B cells and non-B  
 cells  
 Unpublished  
 2 (bases 1 to 596)  
 Martin, A. and Scharff, M.D.  
 Direct Submission  
 Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of  
 Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA  
 Location/Qualifiers  
 1. .596  
 /organism="Homo sapiens"

gene /db\_xref="taxon:9606"  
1. .>596  
/clone="Ramos 5"  
CDS /gene="AID"  
1. .>596  
/gene="AID"  
/note="integrated into Burkitt's lymphoma cell line Ramos"  
/codon\_start=1  
/product="AID"  
/protein\_id="AAM95406.1"  
/db\_xref="GI:22297226"  
/translation="MDSLNNRRKFLYOFKNVRWAKGRRETYLCYVVKRRDSATSFSL  
DFGLRNKNGCHEVLELFLRYISDWLDLPGRYRVWFTSWSPCYDCARHVDLFRGNP  
NLSLRIFTARLYCFDKARPEGLRLRHAGVOIAIMTKDFYCNWTFVENHERTEFK  
AWEGHNSVRLSRQLRILLPLYEVDLDRDARTLGL"  
BASE COUNT 128 a 164 c 154 g 150 t  
ORIGIN

Query Match 19.3%; Score 545; DB 9; Length 596;  
Best Local Similarity 99.8%; Pred. No. 9.8e-273;  
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 80 ATGGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAAATGCCGCTGG 139  
Db 1 ATGGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAAATGCCGCTGG 60  
Qy 140 GCTAAGGGTCGGCTGAGACCTACTCTGTACGTAGTGAAGAGCGCTGCCAGTGGCTACA 199  
Db 61 GCTAAGGGTCGGCTGAGACCTACTCTGTACGTAGTGAAGAGCGCTGCCAGTGGCTACA 120  
Qy 200 TCCTTTTCACTGGACTTTGGTTATCTTCGCAATGAAGACGGCTGCCAGTGGAAATTCCTC 259  
Db 121 TCCTTTTCACTGGACTTTGGTTATCTTCGCAATGAAGACGGCTGCCAGTGGAAATTCCTC 180  
Qy 260 TTCCTCCGTACATCTCGGACTGGGACCTAGACCTGCGCGCTGCTACCGGCTCACCTGG 319  
Db 181 TTCCTCCGTACATCTCGGACTGGGACCTAGACCTGCGCGCTGCTACCGGCTCACCTGG 240  
Qy 320 TTCACCTCCCTGGAGCCCTGCTACGACTGTGCCGAGATGTGGCGGACTTTCTCGGAGGG 379  
Db 241 TTCACCTCCCTGGAGCCCTGCTACGACTGTGCCGAGATGTGGCGGACTTTCTCGGAGGG 300  
Qy 380 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 439  
Db 301 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360  
Qy 440 GCTGAGCCCGAGGGCTGCGGCGCTGCACCGCGCGGGTGCATAAGCATCATGACC 499  
Db 361 GCTGAGCCCGAGGGCTGCGGCGCTGCACCGCGCGGGTGCATAAGCATCATGACC 420  
Qy 500 TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAAGAACTTTCAAA 559  
Db 421 TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAAGAACTTTCAAA 480  
Qy 560 GCCTGGGAAGGGCTGCATGAAATTCAGTTCGTCTCCAGACAGCTTCGGCGCATCCTT 619  
Db 481 GCCTGGGAAGGGCTGCATGAAATTCAGTTCGTCTCCAGACAGCTTCGGCGCATCCTT 540  
Qy 620 TTGCCCTGTATAGGTTGATGACTTACGAGACCGCATTTCTACTTTGGGACTTTG 675  
Db 541 TTGCCCTGTATAGGTTGATGACTTACGAGACCGCATTTCTACTTTGGGACTTTG 596

RESULT 14  
AF529820  
LOCUS AF529820 596 bp mRNA linear PRI 19-AUG-2002  
DEFINITION Homo sapiens clone Ramos 6 AID (AID) mRNA, partial cds.  
ACCESSION AF529820  
VERSION AF529820.1 GI:22297227  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 596)  
AUTHORS Martin,A. and Scharff,M.D.  
TITLE Somatic hypermutation of the AID transgene in B cells and non-B cells  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 596)  
AUTHORS Martin,A. and Scharff,M.D.  
TITLE Direct Submission  
JOURNAL .Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA  
FEATURES  
source  
1. .596  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Ramos 6"  
gene 1. .>596  
/gene="AID"  
CDS 1. .>596  
/gene="AID"  
/note="integrated into Burkitt's lymphoma cell line Ramos"  
/codon\_start=1  
/product="AID"  
/protein\_id="AAM95407.1"  
/db\_xref="GI:22297228"  
/translation="MDSLNNRRKFLYOFKNVRWAKGRRETYLCYVVKRRDSATSFSL  
DFGLRNKNGCHEVLELFLRYISDWLDLPGRYRVWFTSWSPCYDCARHVDLFRGNP  
NLSLRIFTARLYCFDKARPEGLRLRHAGVOIAIMTKDFYCNWTFVENHERTEFK  
AWEGHNSVRLSRQLRILLPLYEVDLDRDARTLGL"  
BASE COUNT 126 a 164 c 155 g 151 t  
ORIGIN  
Query Match 19.3%; Score 545; DB 9; Length 596;  
Best Local Similarity 99.8%; Pred. No. 9.8e-273;  
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 80 ATGGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAAATGCCGCTGG 139  
Db 1 ATGGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAAATGCCGCTGG 60  
Qy 140 GCTAAGGGTCGGCTGAGACCTACTCTGTACGTAGTGAAGAGCGCTGCCAGTGGCTACA 199  
Db 61 GCTAAGGGTCGGCTGAGACCTACTCTGTACGTAGTGAAGAGCGCTGCCAGTGGCTACA 120  
Qy 200 TCCTTTTCACTGGACTTTGGTTATCTTCGCAATGAAGACGGCTGCCAGTGGAAATTCCTC 259  
Db 121 TCCTTTTCACTGGACTTTGGTTATCTTCGCAATGAAGACGGCTGCCAGTGGAAATTCCTC 180  
Qy 260 TTCCTCCGTACATCTCGGACTGGGACCTAGACCTGCGCGCTGCTACCGGCTCACCTGG 319  
Db 181 TTCCTCCGTACATCTCGGACTGGGACCTAGACCTGCGCGCTGCTACCGGCTCACCTGG 240  
Qy 320 TTCACCTCCCTGGAGCCCTGCTACGACTGTGCCGAGATGTGGCGGACTTTCTCGGAGGG 379  
Db 241 TTCACCTCCCTGGAGCCCTGCTACGACTGTGCCGAGATGTGGCGGACTTTCTCGGAGGG 300  
Qy 380 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 439  
Db 301 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360  
Qy 440 GCTGAGCCCGAGGGCTGCGGCGCTGCACCGCGCGGGTGCATAAGCATCATGACC 499  
Db 361 GCTGAGCCCGAGGGCTGCGGCGCTGCACCGCGCGGGTGCATAAGCATCATGACC 420  
Qy 500 TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAAGAACTTTCAAA 559  
Db 421 TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAAGAACTTTCAAA 480  
Qy 560 GCCTGGGAAGGGCTGCATGAAATTCAGTTCGTCTCCAGACAGCTTCGGCGCATCCTT 619  
Db 481 GCCTGGGAAGGGCTGCATGAAATTCAGTTCGTCTCCAGACAGCTTCGGCGCATCCTT 540  
Qy 620 TTGCCCTGTATAGGTTGATGACTTACGAGACCGCATTTCTACTTTGGGACTTTG 675

```
|||||
541 TTGCCCTGTATGAGTTGATGACTTACGAGACGCAATTCGTACTTGGGACTTTG 596

RESULT 15
AF529821
LOCUS Homo sapiens clone Ramos 7 AID (AID) mRNA, partial cds. PRI 19-AUG-2002
DEFINITION
ACCESSION AF529821
VERSION AF529821.1 GI:22297229
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 596)
Martin,A. and Scharff,M.D.
Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
Source
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Ramos 7"
1..>596
/gene="AID"
1..>596
/gene="AID"
/feature="AID"
/feature="Integrated into Burkitt's lymphoma cell line Ramos"
/codon_start=1
/product="AID"
/protein_id="AAM95408.1"
/db_xref="GI:22297230"
/translation="MDSLMMNRKFLYOFKNRWAKGRRETYLCYVVKRRDRAFSFL
DFGLRNKNGCHVELLFLRYISDNDLPDGRGYRTWFTSPCYDCARHVDPLRGNP
NLSLRIFATRIYFCEDRKAPEGLRLHRAVGQIATMTFKDYFCWNTFVENHERTFK
AWELGHENSRLSLRILLPLYEVDLDRAFTLGL"
BASE COUNT 127 a 165 c 155 g 149 t
ORIGIN
Query Match 19.3%; Score 545; DB 9; Length 596;
Best Local Similarity 99.8%; Pred. No. 9.8e-273;
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 80 ATGGACAGCCCTCTTGATGAACCGGAGGAGTTCCTTTACCAATTCAAAAATGTCGCTGG 139
Db 1 ATGGACAGCCCTTGATGAACCGGAGGAGTTCCTTTACCAATTCAAAAATGTCGCTGG 60
Qy 140 GCTAAGGTCGGCTGACACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGTCTACA 199
Db 61 GCTAAGGTCGGCTGACACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGTCTACA 120
Qy 200 TCCCTTTTCACTGGACTTTGGTTATCTCCCAATAAGACGGCTGCCACGTGAATTGCTC 259
Db 121 TCCCTTTTCACTGGACTTTGGTTATCTCCCAATAAGACGGCTGCCACGTGAATTGCTC 180
Qy 260 TTCCTCCGCTACATCTCGACTGGGACCTAGACCCCTGGCCGCTGTCTACCGGCTCACCTGG 319
Db 181 TTCCTCCGCTACATCTCGACTGGGACCCAGACCCCTGGCCGCTGTCTACCGGCTCACCTGG 240
Qy 320 TTCACCTCCTGAGGCCCTGTGTACACTGTGCGCCGACATGTGGCGGCACTTTCTGCGAGGG 379
Db 241 TTCACCTCCTGAGGCCCTGTGTACACTGTGCGCCGACATGTGGCGGCACTTTCTGCGAGGG 300
Qy 380 AACCCCAACCTCAGTCTCAGATCTTACCGCGCCCTCTACTTCTGTGAGGACCGCAAG 439
Db 301 AACCCCAACCTCAGTCTCAGATCTTACCGCGCCCTCTACTTCTGTGAGGACCGCAAG 360
```

```
|||||
440 GCTGAGCCCGAGGGGCTGGCGGCTGCACCGCCCGGGTGCATAATAGCCATCATGACC 499
Db 361 GCTGAGCCCGAGGGGCTGGCGGCTGCACCGCCCGGGTGCATAATAGCCATCATGACC 420
Qy 500 TTCAAAGATTATTTTACTGCTGGAATACATTTCTAGAAAAACCATGAAGAACTTTCAAA 559
Db 421 TTCAAAGATTATTTTACTGCTGGAATACATTTCTAGAAAAACCATGAAGAACTTTCAAA 480
Qy 560 GCCTGGGAAGGCGTGCATGAAAAATTCAGTTCTCTCCAGACAGCTTCGCGCATCCTT 619
Db 481 GCCTGGGAAGGCGTGCATGAAAAATTCAGTTCTCTCCAGACAGCTTCGCGCATCCTT 540
Qy 620 TTGCCCTGTATGAGTTGATGACTTACGAGACGCAATTCGTACTTTGGGACTTTTG 675
Db 541 TTGCCCTGTATGAGTTGATGACTTACGAGACGCAATTCGTACTTTGGGACTTTTG 596

RESULT 16
AF529822
LOCUS Homo sapiens clone Ramos 8 AID (AID) mRNA, partial cds. PRI 19-AUG-2002
DEFINITION
ACCESSION AF529822
VERSION AF529822.1 GI:22297231
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 596)
Martin,A. and Scharff,M.D.
Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
Source
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Ramos 8"
1..>596
/gene="AID"
1..>596
/gene="AID"
/feature="AID"
/feature="Integrated into Burkitt's lymphoma cell line Ramos"
/codon_start=1
/product="AID"
/protein_id="AAM95409.1"
/db_xref="GI:22297232"
/translation="MDSLMMNRKFLYOFKNRWAKGRRETYLCYVVKRRDRAFSFL
DFGLRNKNGCHVELLFLRYISDNDLPDGRGYRTWFTSPCYDCARHVDPLRGNP
NLSLRIFATRIYFCEDRKAPEGLRLHRAVGQIATMTFKDYFCWNTFVENHERTFK
AWELGHENSRLSLRILLPLYEVDLDRAFTLGL"
BASE COUNT 126 a 165 c 155 g 150 t
ORIGIN
Query Match 19.3%; Score 545; DB 9; Length 596;
Best Local Similarity 99.8%; Pred. No. 9.8e-273;
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 80 ATGGACAGCCCTCTTGATGAACCGGAGGAGTTCCTTTACCAATTCAAAAATGTCGCTGG 139
Db 1 ATGGACAGCCCTTGATGAACCGGAGGAGTTCCTTTACCAATTCAAAAATGTCGCTGG 60
Qy 140 GCTAAGGTCGGCTGACACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGTCTACA 199
Db 61 GCTAAGGTCGGCTGACACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGTCTACA 120
Qy 200 TCCCTTTTCACTGGACTTTGGTTATCTCCCAATAAGACGGCTGCCACGTGAATTGCTC 259
```

```
Db 121 TCCCTTTTCTACTGGACTTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGAATTCGCTC 180
QY 260 TTCTCCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCCGCTGCTACCGCGTCACTCGG 319
Db 181 TTCTCCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCCGCTGCTACCGCGTCACTCGG 240
QY 320 TTCACTCTCTGGAGCCCTGCTACGACTGTGCCGACATCTGGCCGACTTTCTTGGCAGGG 379
Db 241 TTCACTCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGGCGACTTTCTTGGCAGGG 300
QY 380 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 439
Db 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360
QY 440 GCTGAGCCCGAGGGGCTGCGCGGCTGCACCGCGCGGGTGCAAATAGCCATCATGACC 499
Db 361 GCTGAGCCCGAGGGGCTGCGCGGCTGCACCGCGCGGGTGCAAATAGCCATCATGACC 420
QY 500 TTCAAGAGATTATTTTACTGCTGGAATCTTTGTAGAAAACCATGAAAGAACTTTTCAAA 559
Db 421 TTCAAGAGATTATTTTACTGCTGGAATCTTTGTAGAAAACCATGAAAGAACTTTTCAAA 480
QY 560 GCTGGAAGGGCTGCATGAATTCAGTTCGCTCTCCAGACAGCTTCGCGCATCTT 619
Db 481 GCTGGAAGGGCTGCATGAATTCAGTTCGCTCTCCAGACAGCTTCGCGCATCTT 540
QY 620 TTGCCCCCTGTATGAGTTGATGACTTACGAGACGCATTTCGACTTTTGGGACTTTG 675
Db 541 TTGCCCCCTGTATGAGTTGATGACTTACGAGACGCATTTCGACTTTTGGGACTTTG 596

RESULT 17
AF529824 Homo sapiens clone Ramos 10 AID (AID) mRNA linear PRI 19-AUG-2002
LOCUS AF529824
DEFINITION Homo sapiens clone Ramos 10 AID (AID) mRNA, partial cds.
ACCESSION AF529824
VERSION AF529824.1 GI:22297235
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
SOURCE 1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Ramos 10"
1..>596
/gene="AID"
1..>596
/gene="AID"
/notes="Integrated into Burkitt's lymphoma cell line Ramos"
/codon_start=1
/product="AID"
/protein_id="AAM95411.1"
/db_xref="GI:22297236"
/translocation="MDSLLNMRKKFLYQKFNVRWAKGRRETYLCYVVKRRDSATSFSL
DFGYLRNKGWVEVLFLRYISDWLDPGRCYRVTFWTSYPCYCARHVADEFLRGNP
NLSLRIFFARLYFCEDRKAEPGLRRLRHAGVQIAIMTEKDYFCWNTVEVHERTFK
AWEGLHENSVRLSROLRRLILLPLYEVDLDRDAFRTLGL"
BASE COUNT 127 a 163 c 156 g 150 t
ORIGIN
```

```
Query Match 19.3%; Score 545; DB 9; Length 596;
Best Local Similarity 99.8%; Pred. No. 9.8e-273;
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 80 ATGGACAGCCTCTTGATGAACCGGAGGAAGTTTCTTTACCAATTCAAAAATGTCCGCTGG 139
Db 1 ATGGACAGCCTCTTGATGAACCGGAGGAAGTTTCTTTACCAATTCAAAAATGTCCGCTGG 60
QY 140 GCTAAGGTCGCGGTGAGACCTACCTGTGCTAGTAGTGAAGAGCGGTGACAGTGTCTACA 199
Db 61 GCTAAGGTCGCGGTGAGACCTACCTGTGCTAGTAGTGAAGAGCGGTGACAGTGTCTACA 120
QY 200 TCTCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGAATTCGCTC 259
Db 121 TCTCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGAATTCGCTC 180
QY 260 TTCTCCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCCGCTGCTACCGCGTCACTCGG 319
Db 181 TTCTCCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCCGCTGCTACCGCGTCACTCGG 240
QY 320 TTCACTCTCTGGAGCCCTGCTACGACTGTGCCGACATCTGGCCGACTTTCTTGGCAGGG 379
Db 241 TTCACTCTCTGGAGCCCTGCTACGACTGTGCCGACATCTGGCCGACTTTCTTGGCAGGG 300
QY 380 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 439
Db 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360
QY 440 GCTGAGCCCGAGGGGCTGCGCGGCTGCACCGCGCGGGTGCAAATAGCCATCATGACC 499
Db 361 GCTGAGCCCGAGGGGCTGCGCGGCTGCACCGCGCGGGTGCAAATAGCCATCATGACC 420
QY 500 TTCAAGAGATTATTTTACTGCTGGAATCTTTGTAGAAAACCATGAAAGAACTTTTCAAA 559
Db 421 TTCAAGAGATTATTTTACTGCTGGAATCTTTGTAGAAAACCATGAAAGAACTTTTCAAA 480
QY 560 GCTTGGGAAGGGCTGCATGAATTCAGTTCGCTCTCCAGACAGCTTCGCGCATCTT 619
Db 481 GCTTGGGAAGGGCTGCATGAATTCAGTTCGCTCTCCAGACAGCTTCGCGCATCTT 540
QY 620 TTGCCCCCTGTATGAGTTGATGACTTACGAGACGCATTTCGACTTTTGGGACTTTG 675
Db 541 TTGCCCCCTGTATGAGTTGATGACTTACGAGACGCATTTCGACTTTTGGGACTTTG 596

RESULT 18
AF529825 Homo sapiens clone Ramos 11 AID (AID) mRNA linear PRI 19-AUG-2002
LOCUS AF529825
DEFINITION Homo sapiens clone Ramos 11 AID (AID) mRNA, partial cds.
ACCESSION AF529825
VERSION AF529825.1 GI:22297237
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
SOURCE 1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Ramos 11"
1..>596
/gene="AID"
1..>596
/gene="AID"
/notes="Integrated into Burkitt's lymphoma cell line Ramos"
/codon_start=1
/product="AID"
/protein_id="AAM95411.1"
/db_xref="GI:22297236"
/translocation="MDSLLNMRKKFLYQKFNVRWAKGRRETYLCYVVKRRDSATSFSL
DFGYLRNKGWVEVLFLRYISDWLDPGRCYRVTFWTSYPCYCARHVADEFLRGNP
NLSLRIFFARLYFCEDRKAEPGLRRLRHAGVQIAIMTEKDYFCWNTVEVHERTFK
AWEGLHENSVRLSROLRRLILLPLYEVDLDRDAFRTLGL"
BASE COUNT 127 a 163 c 156 g 150 t
ORIGIN
```

gene

```

CDS
/feature="AID"
1..>596
/feature="AID"
/feature="integrated into Burkitt's lymphoma cell line Ramos"
/codon_start=1
/product="AID"
/protein_id="AAM95412.1"
/db_xref="GI:22297238"
/translation="MDSLIMNRKFLYQFNVRWAKGRREAYLCYVVKRRDSATSFSL
DFGLRNKNGCHVELLFRLYISDWLDLPGRCYRVTWFTSWSPCYDCARHVAFLRGNP
NLSLRFTARLYFCEDRKAEPGLRLRHAGVQIAIMTFKDYFCWNTFVENHIERFK
AWEGLHENSVRLSRQLRILLPLYEVDLDRDAFRTLGL"
BASE COUNT 126 a 164 c 156 g 150 t
ORIGIN
Query Match 19.3%; Score 545; DB 9; Length 596;
Best Local Similarity 99.8%; Pred. No. 9.8e-273;
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 80 ATGCACAGCCTCTTGATGAACCGAGGAAAGTTCTTTACCAATTCAAAAATGTCGCTGG 139
Db 1 ATGCACAGCCTCTTGATGAACCGAGGAAAGTTCTTTACCAATTCAAAAATGTCGCTGG 60
QY 140 GCTAAGGGTCGGCTGAGACCTACCTGTGCTAGTGAAGAGGGCTACACAGTGTACACA 199
Db 61 GCTAAGGGTCGGCTGAGACCTACCTGTGCTAGTGAAGAGGGCTACACAGTGTACACA 120
QY 200 TCCTTTTCACTGGACCTTGGTTATCTTCGCAATGAAGACGGCTGCCACGTGGAAATGGCTC 259
Db 121 TCCTTTTCACTGGACCTTGGTTATCTTCGCAATGAAGACGGCTGCCACGTGGAAATGGCTC 180
QY 260 TTCTCGGCTACATCTCGGACTGGACCTAGACCCCTGGCGGCTGCTACCCGCTCACTGG 240
Db 181 TTCTCGGCTACATCTCGGACTGGACCTAGACCCCTGGCGGCTGCTACCCGCTCACTGG 379
QY 320 TTCACTCTCTGGAGCCCTGCTAGGACTGTGCCGGACATGTGGCCGACTTTCTGCGAGGG 300
Db 241 TTCACTCTCTGGAGCCCTGCTAGGACTGTGCCGGACATGTGGCCGACTTTCTGCGAGGG 439
QY 380 AACCCCAACCTCAGTCTGAGGATCTTCCACCGCGGCTGCAAAATAGCCATCATGACC 499
Db 301 AACCCCAACCTCAGTCTGAGGATCTTCCACCGCGGCTGCAAAATAGCCATCATGACC 420
QY 440 GCTGAGCCCGAGGGCTGCGGCGCTGACCGCGCGGCTGCAAAATAGCCATCATGACC 559
Db 361 GCTGAGCCCGAGGGCTGCGGCGCTGACCGCGCGGCTGCAAAATAGCCATCATGACC 480
QY 500 TTCAAAAGATTATTTTACTGCTGGAATACCTTTGTTAGAAAACCATGAAAGAACTTTTCAAA 675
Db 421 TTCAAAAGATTATTTTACTGCTGGAATACCTTTGTTAGAAAACCATGAAAGAACTTTTCAAA 596
QY 560 GCCTGGGAAGGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTCGCGCATCCTT 619
Db 481 GCCTGGGAAGGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTCGCGCATCCTT 540
QY 620 TTGCCCCCTGTATGAGGTTGATGACTTACGAGACGCCATTTGCTACTTTGGGACTTTG 675
Db 541 TTGCCCCCTGTATGAGGTTGATGACTTACGAGACGCCATTTGCTACTTTGGGACTTTG 596
RESULT 19
AF529826 596 bp mRNA linear PRI 19-AUG-2002
LOCUS Homo sapiens clone Ramos 12 AID (AID) mRNA, partial cds.
DEFINITION AF529826
ACCESSION AF529826
VERSION AF529826.1 GI:22297239
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
```

```

TITLE Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Ramos 12"
1..>596
/feature="AID"
1..>596
/feature="AID"
/feature="integrated into Burkitt's lymphoma cell line Ramos"
/codon_start=1
/product="AID"
/protein_id="AAM95413.1"
/db_xref="GI:22297240"
/translation="MDSLIMNRKFLYQFNVRWAKGRRETYLCYVVKRRDSATSFSL
DFGLRNKNGCHVELLFRLYISDWLDLPGRCYRVTWFTSWSPCYDCARHVAFLRGNP
NLSLRFTARLYFCEDRKAEPGLRLRHAGVQIAIMTFKDYFCWNTFVENHIERFK
AWEGLHENSVRLSRQLRILLPLYEVDLDRDAFRTLGL"
BASE COUNT 127 a 163 c 155 g 151 t
ORIGIN
Query Match 19.3%; Score 545; DB 9; Length 596;
Best Local Similarity 99.8%; Pred. No. 9.8e-273;
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 80 ATGCACAGCCTCTTGATGAACCGAGGAAAGTTCTTTACCAATTCAAAAATGTCGCTGG 139
Db 1 ATGCACAGCCTCTTGATGAACCGAGGAAAGTTCTTTACCAATTCAAAAATGTCGCTGG 60
QY 140 GCTAAGGGTCGGCTGAGACCTACCTGTGCTAGTGAAGAGGGCTACACAGTGTACACA 199
Db 61 GCTAAGGGTCGGCTGAGACCTACCTGTGCTAGTGAAGAGGGCTACACAGTGTACACA 120
QY 200 TCCTTTTCACTGGACCTTGGTTATCTTCGCAATGAAGACGGCTGCCACGTGGAAATGGCTC 259
Db 121 TCCTTTTCACTGGACCTTGGTTATCTTCGCAATGAAGACGGCTGCCACGTGGAAATGGCTC 180
QY 260 TTCTCGGCTACATCTCGGACTGGACCTAGACCCCTGGCGGCTGCTACCCGCTCACTGG 319
Db 181 TTCTCGGCTACATCTCGGACTGGACCTAGACCCCTGGCGGCTGCTACCCGCTCACTGG 240
QY 320 TTCACTCTCTGGAGCCCTGCTAGGACTGTGCCGGACATGTGGCCGACTTTCTGCGAGGG 379
Db 241 TTCACTCTCTGGAGCCCTGCTAGGACTGTGCCGGACATGTGGCCGACTTTCTGCGAGGG 300
QY 380 AACCCCAACCTCAGTCTGAGGATCTTCCACCGCGGCTGCAAAATAGCCATCATGACC 419
Db 301 AACCCCAACCTCAGTCTGAGGATCTTCCACCGCGGCTGCAAAATAGCCATCATGACC 360
QY 440 GCTGAGCCCGAGGGCTGCGGCGCTGACCGCGCGGCTGCAAAATAGCCATCATGACC 499
Db 361 GCTGAGCCCGAGGGCTGCGGCGCTGACCGCGCGGCTGCAAAATAGCCATCATGACC 420
QY 500 TTCAAAAGATTATTTTACTGCTGGAATACCTTTTGTAGAAAACCATGAAAGAACTTTTCAAA 559
Db 421 TTCAAAAGATTATTTTACTGCTGGAATACCTTTTGTAGAAAACCATGAAAGAACTTTTCAAA 480
QY 560 GCCTGGGAAGGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTCGCGCATCCTT 619
Db 481 GCCTGGGAAGGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTCGCGCATCCTT 540
QY 620 TTGCCCCCTGTATGAGGTTGATGACTTACGAGACGCCATTTGCTACTTTGGGACTTTG 675
Db 541 TTGCCCCCTGTATGAGGTTGATGACTTACGAGACGCCATTTGCTACTTTGGGACTTTG 596
```

```
RESULT 20
AF529827
LOCUS Homo sapiens clone Ramos 13 AID (AID) mRNA, partial cds.
DEFINITION AF529827
ACCESSION AF529827
VERSION AF529827.1 GI:22297241
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 596)
AUTHORS Martin.A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin.A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
Location/Qualifiers
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Ramos 13"
1..>596
/gene="AID"
1..>596
/gene="AID"
/note="Integrated into Burkitt's lymphoma cell line Ramos"
/codon_start=1
/product="AID"
/protein_id="AAM95414.1"
/db_xref="GI:22297242"
/translation="MDSLMMRRKFLYQFKVNRWAKGRRETYLCVWVRRSATSFSL
DFGLRNKNGCHVELLFYISDWLDPRCYRVTFWTSWSPCYCAHRVADFLRGNP
NLSRIFARLYFCEDRKAEPGLRRLHRAGVQIAIMTFKDYFCWNTFVENHRTFFK
AWEGLHNSVRLSRLRLPLLYEVDLROAFRTLGL"
BASE COUNT 127 a 163 c 155 g 151 t
ORIGIN
Query Match 19.3%; Score 545; DB 9; Length 596;
Best Local Similarity 99.8%; Pred. No. 9.8e-273;
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 80 ATGGACAGCCTCTTGATCAACCGGAGGAGTTTCTTTACCAATTCAAAAATGTCGCTGG 139
Db 1 ATGGACAGCCTCTTGATCAACCGGAGGAGTTTCTTTACCAATTCAAAAATGTCGCTGG 60
QY 140 GCTAAGGTCGCGGTGAGACCTTACCTGTGCTACGTAGTGAAGCGGTGACAGTGTCTACA 199
Db 61 GCTAAGGTCGCGGTGAGACCTTACCTGTGCTACGTAGTGAAGCGGTGACAGTGTCTACA 120
QY 200 TCCTTTTCTACGTGAGCTTTGGTTATCTTCGCAATAGACGGCTGCCAGTGAANTGTCTC 259
Db 121 TCCTTTTCTACGTGAGCTTTGGTTATCTTCGCAATAGACGGCTGCCAGTGAANTGTCTC 180
QY 260 TTCTTCCGCTACATCTCGGACTGGGACCTAGACCTGCGCGCTGCTACCGGCTCACTGG 319
Db 181 TTCTTCCGCTACATCTCGGACTGGGACCTAGACCTGCGCGCTGCTACCGGCTCACTGG 240
QY 320 TTCACTTCTTGAGGCCCTCTGCTACGACTGTGCGCGACATGTGCGCGACTTTCTGCGAGGG 379
Db 241 TTCACTTCTTGAGGCCCTCTGCTACGACTGTGCGCGACATGTGCGCGACTTTCTGCGAGGG 300
QY 380 ACCCCCAACCTCAGTCTGAGGATCTTACCGCGCGCCCTCTACTTCTGTGAGAGCCGCAAG 439
Db 301 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCCCTCTACTTCTGTGAGAGCCGCAAG 360
QY 440 GCTGAGCCCGAGGGCTCGCGCGCTGCACCGCGGGGTGCAATAGCCATCATGACC 499
|||||
```

```
Db 361 GCTGAGCCCGAGGGCTCGCGCGCTGCACCGCGGGTGCAAAATAGCCATCATGACC 420
QY 500 TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAAGAACTTTCAA 559
|||||
Db 421 TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAAGAACTTTCAA 480
QY 560 GCCTCGGAAGGGCTGCATGAAATTCAGTTCGTCCTCCAGACAGCTTCGGCGCATCCPT 619
Db 481 GCCTCGGAAGGGCTGCATGAAATTCAGTTCGTCCTCCAGACAGCTTCGGCGCATCCPT 540
QY 620 TTGCCCCCTGTATGAGGTTGATGACATTACGAGACGATTTTCGTACTTTGGGACTTTG 675
Db 541 TTGCCCCCTGTATGAGGTTGATGACATTACGAGACGATTTTCGTACTTTGGGACTTTG 596
RESULT 21
AF529830
LOCUS Mus musculus clone 3 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION complete cds.
ACCESSION AF529830
VERSION AF529830
KEYWORDS house mouse.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 596)
AUTHORS Martin.A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B and non-B cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin.A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
Location/Qualifiers
1..596
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="3"
/cell_type="hybridoma P1-5"
/transgenic
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..>596
/gene="AID"
1..>596
/gene="AID"
/codon_start=1
/product="AID"
/protein_id="AAM95417.1"
/db_xref="GI:22297248"
/translation="MDSLMMRRKFLYQFKVNRWAKGRRETYLCVWVRRSATSFSL
DFGLRNKNGCHVELLFYISDWLDPRCYRVTFWTSWSPCYCAHRVADFLRGNP
NLSRIFARLYFCEDRKAEPGLRRLHRAGVQIAIMSFKDYFCWNTFVENHRTFFK
AWEGLHNSVRLSRLRLPLLYEVDLROAFRTLGL"
BASE COUNT 126 a 164 c 155 g 151 t
ORIGIN
Query Match 19.3%; Score 545; DB 12; Length 596;
Best Local Similarity 99.8%; Pred. No. 9.8e-273;
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 80 ATGGACAGCCTCTTGATCAACCGGAGGAGTTTCTTTACCAATTCAAAAATGTCGCTGG 139
Db 1 ATGGACAGCCTCTTGATCAACCGGAGGAGTTTCTTTACCAATTCAAAAATGTCGCTGG 60
QY 140 GCTAAGGTCGCGGTGAGACCTTACCTGTGCTACGTAGTGAAGCGGTGACAGTGTCTACA 199
Db 61 GCTAAGGTCGCGGTGAGACCTTACCTGTGCTACGTAGTGAAGCGGTGACAGTGTCTACA 120
```

```
QY 200 TCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGAACGGCTGCCAGTGGAAATTCGTC 259
Db 121 TCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGAACGGCTGCCAGTGGAAATTCGTC 180
QY 260 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCCGCTGCTACCGGCTCACCTGG 319
Db 181 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCCGCTGCTACCGGCTCACCTGG 240
QY 320 TTCACCTCCTGGAGCCCTGCTAGACATGTGCCGACATGTGCCGACATGTCTCGAGGG 379
Db 241 TTCACCTCCTGGAGCCCTGCTAGACATGTGCCGACATGTGCCGACATGTCTCGAGGG 300
QY 380 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCCCTCTACTTCTGTGAGGACCGCAAG 439
Db 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCCCTCTACTTCTGTGAGGACCGCAAG 360
QY 440 GCTGAGCCCGAGGGCTCGCGGGCTGCACCGCGCGGGGTGCAAAATAGCCATCATGACC 499
Db 361 GCTGAGCCCGAGGGCTCGCGGGCTGCACCGCGCGGGGTGCAAAATAGCCATCATGACC 420
QY 500 TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAACCATGAAGAACTTTTCAA 559
Db 421 TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAACCATGAAGAACTTTTCAA 480
QY 560 GCCTGGGAGGGCTGCATGAAATTCAGTTCGTCTCTCCAGACAGCTTCGCGCGCATCCTT 619
Db 481 GCCTGGGAGGGCTGCATGAAATTCAGTTCGTCTCTCCAGACAGCTTCGCGCGCATCCTT 540
QY 620 TTGCCCTGTATGAGTTGATGACTTAGAGACGCAATTCGTACTTTGGGACTTTG 675
Db 541 TTGCCCTGTATGAGTTGATGACTTAGAGACGCAATTCGTACTTTGGGACTTTG 596

RESULT 22
AF529831
LOCUS AF529831 Mus musculus clone 4 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION complete cds.
ACCESSION AF529831
VERSION AF529831.1 GI:22297249
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE Martin,A. and Scharff,M.D.
JOURNAL Somatic hypermutation of the AID transgene in B and non-B cells
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source
1..596
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="4"
/cell_type="hybridoma P1-5"
/transgenic
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/gene="AID"
1..>596
/gene="AID"
1..>596
/codon_start=1
/product="AID"
/protein_id="AAM95418.1"
/db_xref="GI:22297250"
/translation="MDSLMMNRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSL
```

```
DFGYLRNKGCHVELLEFLRYISDWDLDPGRCYRVMTFTSWSPCYDCAKHVDFLRGNP
NLSRIETARLYFCEDRKAEPEGLRHAGVQIAIMTFKDYFYCWNTFVENIERTEK
AWEGLHNSVRLSRQFRILLPLLYEVDLRDAPRTIGL"
BASE COUNT 127 a 163 c 155 g 151 t
ORIGIN
Query Match 19.3%; Score 545; DB 12; Length 596;
Best Local Similarity 99.8%; Pred. No. 9.8e-273;
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 80 ATGACAGCCCTCTTGATGAAACGGGAGGAAGTTTCTTTACCAATTCACAAATTCCTCGCTGG 139
Db 1 ATGACAGCCCTCTTGATGAAACGGGAGGAAGTTTCTTTACCAATTCACAAATTCCTCGCTGG 60
QY 140 GCTAAGGCTCGGCTGAGACCTACCTGCTGTACCTAGTGAAGAGCGGTGACAGTGCTACA 199
Db 61 GCTAAGGCTCGGCTGAGACCTACCTGCTGTACCTAGTGAAGAGCGGTGACAGTGCTACA 120
QY 200 TCCTTTTCACTGGACTTTTGGTTATCTTCGCAATAAGAACGGCTGCCAGTGGAAATTCGTC 259
Db 121 TCCTTTTCACTGGACTTTTGGTTATCTTCGCAATAAGAACGGCTGCCAGTGGAAATTCGTC 180
QY 260 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCCGCTGCTACCGGCTCACCTGG 319
Db 181 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCCGCTGCTACCGGCTCACCTGG 240
QY 320 TTCACCTCCTGGAGCCCTGCTAGACATGTGCCGACATGTGCCGAGCTTCTCGGAGGG 379
Db 241 TTCACCTCCTGGAGCCCTGCTAGACATGTGCCGACATGTGCCGAGCTTCTCGGAGGG 300
QY 380 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCCCTCTACTTCTGTGAGGACCGCAAG 439
Db 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCCCTCTACTTCTGTGAGGACCGCAAG 360
QY 440 GCTGAGCCCGAGGGCTCGCGGGCTGCACCGCGCGGGGTGCAAAATAGCCATCATGACC 499
Db 361 GCTGAGCCCGAGGGCTCGCGGGCTGCACCGCGCGGGGTGCAAAATAGCCATCATGACC 420
QY 500 TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAACCATGAAGAACTTTTCAA 559
Db 421 TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAACCATGAAGAACTTTTCAA 480
QY 560 GCCTGGGAGGGCTGCATGAAATTCAGTTCGTCTCTCCAGACAGCTTCGCGGCGCATCTT 619
Db 481 GCCTGGGAGGGCTGCATGAAATTCAGTTCGTCTCTCCAGACAGCTTCGCGGCGCATCTT 540
QY 620 TTGCCCTGTATGAGTTGATGACTTAGAGACGCAATTCGTACTTTGGGACTTTG 675
Db 541 TTGCCCTGTATGAGTTGATGACTTAGAGACGCAATTCGTACTTTGGGACTTTG 596

RESULT 23
AF529833
LOCUS AF529833 Mus musculus clone 6 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION complete cds.
ACCESSION AF529833
VERSION AF529833.1 GI:22297253
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE Martin,A. and Scharff,M.D.
JOURNAL Somatic hypermutation of the AID transgene in B and non-B cells
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
Location/Qualifiers
```



```
source
1. .596
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="6"
/cell_type="hybridoma P1-5"
/transgenic
source
1. .596
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="8"
/cell_type="hybridoma P1-5"
/transgenic
1. .596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="8"
/genes="AID"
/genes="AID"
/codon_start=1
/product="AID"
/protein_id="AAM95420.1"
/db_xref="GI:22297254"
/translation="MDSLNNRRKFLYQKNVWAKGRRETYVCYVVKRDSATSFSL
DFGRLNKGCHVELLFLRYISDWLDLPGRCYRVWFTSWSPCYDCARHVADFLRNP
NLSRIFTARLYFCDKAEGLRLHRAVQIAIMTKDYFCWNTFVENHERTFK
AWGLHENSRLSRQLRLLPLYEVDLRLDAFRTLGL"
BASE COUNT 127 a 163 c 156 g 150 t
ORIGIN
Query Match 19.3%; Score 545; DB 12; Length 596;
Best Local Similarity 99.8%; Pred. No. 9.8e-273;
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 80 ATGCACAGCCTCTTGATGAACCGGAGGAGTCTTTTACCAATTCAAAAATGCCGTGG 139
Db 1 ATGCACAGCCTCTTGATGAACCGGAGGAGTCTTTTACCAATTCAAAAATGCCGTGG 60
Qy 140 GCTAAGGGTCGGGTGAGACCTACCTGTGTAGTGTAGTGAAGAGGGGTGACAGTGTCTA 199
Db 61 GCTAAGGGTCGGGTGAGACCTACCTGTGTAGTGTAGTGAAGAGGGGTGACAGTGTCTA 120
Qy 200 TCCTTTTCACTGACATTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGGAATGCTC 259
Db 121 TCCTTTTCACTGACATTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGGAATGCTC 180
Qy 260 TTCCTCGCTACATCTCGGACTGGGACCTAGACCCCTGGCGGCTGCTACCTGCTCACCTGG 319
Db 181 TTCCTCGCTACATCTCGGACTGGGACCTAGACCCCTGGCGGCTGCTACCTGCTCACCTGG 240
Qy 320 TTCACCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGCCGCGACTTTCTCGCAGGG 379
Db 241 TTCACCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGCCGCGACTTTCTCGCAGGG 300
Qy 380 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCTCTACTCTGTGAGGACCGCAAG 439
Db 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCTCTACTCTGTGAGGACCGCAAG 360
Qy 440 GCTGAGCCCGAGGGGTGCGGCGCTGCACCGCGGGGTGCAAAATAGCATCATGACC 499
Db 361 GCTGAGCCCGAGGGGTGCGGCGCTGCACCGCGGGGTGCAAAATAGCATCATGACC 420
Qy 500 TTCAAAGATTATTTTACTGCTGGAATCTTTGTAGAAACCATGAAGAACTTTCAA 559
Db 421 TTCAAAGATTATTTTACTGCTGGAATCTTTGTAGAAACCATGAAGAACTTTCAA 480
Qy 560 GCCTGGGAAGGGCTGCATGAATTCAGTTTCGTCTCTCCAGACAGCTTCGGCGCATCCTT 619
Db 481 GCCTGGGAAGGGCTGCATGAATTCAGTTTCGTCTCTCCAGACAGCTTCGGCGCATCCTT 540
Qy 620 TTGCCCTCTGATGAGGTTGATGACTTACGAGACGCAATTCGTACTTTGGGACTTTG 675
Db 541 TTGCCCTCTGATGAGGTTGATGACTTACGAGACGCAATTCGTACTTTGGGACTTTG 596
RESULT 24
AF529835 596 bp mRNA linear SYN 19-AUG-2002
LOCUS Mus musculus clone 8 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION
```

```
complete cds.
ACCESSION AF529835
VERSION AF529835.1 GI:22297257
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 596)
AUTHORS Martin, A. and Scharff, M.D.
TITLE Somatic hypermutation of the AID transgene in B and non-B cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin, A. and Scharff, M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source
1. .596
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="8"
/cell_type="hybridoma P1-5"
/transgenic
1. .596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="8"
/genes="AID"
/genes="AID"
/codon_start=1
/product="AID"
/protein_id="AAM95422.1"
/db_xref="GI:22297258"
/translation="MDSLNNRRKFLYQKNVWAKGRRETYVCYVVKRDSATSFSL
DFGRLNKGCHVELLFLRYISDWLDLPGRCYRVWFTSWSPCYDCARHVADFLRNP
NLSRIFTARLYFCDKAEGLRLHRAVQIAIMTKDYFCWNTFVENHERTFK
AWGLHENSRLSRQLRLLPLYEVDLRLDAFRTLGL"
BASE COUNT 127 a 163 c 156 g 150 t
ORIGIN
Query Match 19.3%; Score 545; DB 12; Length 596;
Best Local Similarity 99.8%; Pred. No. 9.8e-273;
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 80 ATGCACAGCCTCTTGATGAACCGGAGGAGTCTTTTACCAATTCAAAAATGCCGTGG 139
Db 1 ATGCACAGCCTCTTGATGAACCGGAGGAGTCTTTTACCAATTCAAAAATGCCGTGG 60
Qy 140 GCTAAGGGTCGGGTGAGACCTACCTGTGTAGTGTAGTGAAGAGGGGTGACAGTGTCTA 199
Db 61 GCTAAGGGTCGGGTGAGACCTACCTGTGTAGTGTAGTGAAGAGGGGTGACAGTGTCTA 120
Qy 200 TCCTTTTCACTGACATTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGGAATGCTC 259
Db 121 TCCTTTTCACTGACATTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGGAATGCTC 180
Qy 260 TTCCTCGCTACATCTCGGACTGGGACCTAGACCCCTGGCGGCTGCTACCTGCTCACCTGG 319
Db 181 TTCCTCGCTACATCTCGGACTGGGACCTAGACCCCTGGCGGCTGCTACCTGCTCACCTGG 240
Qy 320 TTCACCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGCCGCGACTTTCTCGCAGGG 379
Db 241 TTCACCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGCCGCGACTTTCTCGCAGGG 300
Qy 380 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCTCTACTCTGTGAGGACCGCAAG 439
Db 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCTCTACTCTGTGAGGACCGCAAG 360
Qy 440 GCTGAGCCCGAGGGGTGCGGCGCTGCACCGCGGGGTGCAAAATAGCATCATGACC 499
Db 361 GCTGAGCCCGAGGGGTGCGGCGCTGCACCGCGGGGTGCAAAATAGCATCATGACC 420
Qy 500 TTCAAAGATTATTTTACTGCTGGAATCTTTGTAGAAACCATGAAGAACTTTCAA 559
Db 421 TTCAAAGATTATTTTACTGCTGGAATCTTTGTAGAAACCATGAAGAACTTTCAA 480
Qy 560 GCCTGGGAAGGGCTGCATGAATTCAGTTTCGTCTCTCCAGACAGCTTCGGCGCATCCTT 619
Db 481 GCCTGGGAAGGGCTGCATGAATTCAGTTTCGTCTCTCCAGACAGCTTCGGCGCATCCTT 540
Qy 620 TTGCCCTCTGATGAGGTTGATGACTTACGAGACGCAATTCGTACTTTGGGACTTTG 675
Db 541 TTGCCCTCTGATGAGGTTGATGACTTACGAGACGCAATTCGTACTTTGGGACTTTG 596
RESULT 24
AF529835 596 bp mRNA linear SYN 19-AUG-2002
LOCUS Mus musculus clone 8 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION
```



BASE COUNT 128 a 163 c 155 g 150 t  
ORIGIN  
NLSRIFTARLYFCBDRKAETEGRLRLHRAGVQIAIMTFKDYFYCWNTFVENHRTFK  
AWEGHLHNSVRLSRILLPLLYEVDDLRAFRTLGL"

Query Match 19.3%; Score 545; DB 12; Length 596;  
Best Local Similarity 99.8%; Pred. No. 9.8e-273;  
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 80 ATGGACAGCCTCTTGATGAACCGGAGGAGTTTCTTTACCAATTCAAAAATCGCGCTGG 139  
Db 1 ATGGACAGCCTCTTGATGAACCGGAGGAGTTTCTTTACCAATTCAAAAATCGCGCTGG 60  
QY 140 GCTAAGGGTGGCGTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGTCTACA 199  
Db 61 GCTAAGGGTGGCGTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGTCTACA 120  
QY 200 TCTTTTCTACTGGACTTTGGTTATCTTCGCAATAGACGGTGCACGTGGAATTTGCTC 259  
Db 121 TCTTTTCTACTGGACTTTGGTTATCTTCGCAATAGACGGTGCACGTGGAATTTGCTC 180  
QY 260 TTCTCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCGGCTGCTACCGGCTACCTGG 319  
Db 181 TTCTCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCGGCTGCTACCGGCTACCTGG 240  
QY 320 TTACCTCTCTGGAGCCCTGCTACGACTGTGCGCGACATGTGGCGGACTTTCTGGGAGGG 379  
Db 241 TTACCTCTCTGGAGCCCTGCTACGACTGTGCGCGACATGTGGCGGACTTTCTGGGAGGG 300  
QY 380 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 439  
Db 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360  
QY 440 GCTGAGCCGAGGGCTGGCGGCTGCACCGCGCGGGTGCAATAGCCATCATGACC 499  
Db 361 GCTGAGCCGAGGGCTGGCGGCTGCACCGCGCGGGTGCAATAGCCATCATGACC 420  
QY 500 TTCAAGATATTTTTACTGCTGGAATCTTTTGTAGAAAACCATGAAGAACTTTTCAA 559  
Db 421 TTCAAGATATTTTTACTGCTGGAATCTTTTGTAGAAAACCATGAAGAACTTTTCAA 480  
QY 560 GCTGGGAAGGCTGCATGAATAATTCAGTTCGCTCTCCAGACAGCTTCGGCGCATCTTT 619  
Db 481 GCTGGGAAGGCTGCATGAATAATTCAGTTCGCTCTCCAGACAGCTTCGGCGCATCTTT 540  
QY 620 TTGCCCCCTGATGAGGTTGATGACTTACGAGACGCAATTCGTACTTTGGGACTTTG 675  
Db 541 TTGCCCCCTGATGAGGTTGATGACTTACGAGACGCAATTCGTACTTTGGGACTTTG 596

RESULT 27  
AF529834  
LOCUS  
DEFINITION Mus musculus clone 7 transgenic Homo sapiens AID (AID) mRNA,  
complete cds.  
ACCESSION  
VERSION AF529834.1 GI:22297255  
KEYWORDS  
SOURCE house mouse.  
ORGANISM Mus musculus  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 595)  
AUTHORS Martin,A. and Scharff,M.D.  
TITLE Somatic hypermutation of the AID transgene in B and non-B cells  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 595)  
AUTHORS Martin,A. and Scharff,M.D.  
TITLE Direct Submission  
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of  
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA  
FEATURES  
source Location/Qualifiers  
1..595

/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="7"  
/cell\_type="hybridoma P1-5"  
/transgenic  
1..595  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
1..>595  
/gene="AID"  
/gene="AID"  
1..>595  
/codon\_start=1  
/product="AID"  
/protein\_id="AA095421.1"  
/db\_xref="GI:22297256"  
/translation="MDSLLMNRKFLYQKVRNKAQRRETYLCVIVKRRSATSFSLS  
DFGLRNKNGCHVELLFURYISDWDLPGRYRVTFWTSWPCYDCAHRVADFLRGNP  
NLSRIFTARLYFCBDRKAETEGRLRLHRAGVQIAIMTFKDYFYCWNTFVENHRTFK  
AWEGHLHNSVRLSRILLPLLYEVDDLRAFRTLGL"  
BASE COUNT 127 a 163 c 154 g 151 t  
ORIGIN

Query Match 19.3%; Score 544; DB 12; Length 595;  
Best Local Similarity 99.8%; Pred. No. 3.3e-272;  
Matches 594; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 80 ATGGACAGCCTCTTGATGAACCGGAGGAGTTTCTTTACCAATTCAAAAATCGCGCTGG 139  
Db 1 ATGGACAGCCTCTTGATGAACCGGAGGAGTTTCTTTACCAATTCAAAAATCGCGCTGG 60  
QY 140 GCTAAGGGTGGCGTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGTCTACA 199  
Db 61 GCTAAGGGTGGCGTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGTCTACA 120  
QY 200 TCTTTTCTACTGGACTTTGGTTATCTTCGCAATAGACGGTGCACGTGGAATTTGCTC 259  
Db 121 TCTTTTCTACTGGACTTTGGTTATCTTCGCAATAGACGGTGCACGTGGAATTTGCTC 180  
QY 260 TTCTCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCGGCTGCTACCGGCTACCTGG 319  
Db 181 TTCTCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCGGCTGCTACCGGCTACCTGG 240  
QY 320 TTACCTCTCTGGAGCCCTGCTACGACTGTGCGCGACATGTGGCGGACTTTCTGGGAGGG 379  
Db 241 TTACCTCTCTGGAGCCCTGCTACGACTGTGCGCGACATGTGGCGGACTTTCTGGGAGGG 300  
QY 380 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 439  
Db 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360  
QY 440 GCTGAGCCGAGGGCTGGCGGCTGCACCGCGCGGGTGCAATAGCCATCATGACC 499  
Db 361 GCTGAGCCGAGGGCTGGCGGCTGCACCGCGCGGGTGCAATAGCCATCATGACC 420  
QY 500 TTCAAGATATTTTTACTGCTGGAATCTTTTGTAGAAAACCATGAAGAACTTTTCAA 559  
Db 421 TTCAAGATATTTTTACTGCTGGAATATTTTGTAGAAAACCATGAAGAACTTTTCAA 480  
QY 560 GCTGGGAAGGCTGCATGAATAATTCAGTTCGCTCTCCAGACAGCTTCGGCGCATCTTT 619  
Db 481 GCTGGGAAGGCTGCATGAATAATTCAGTTCGCTCTCCAGACAGCTTCGGCGCATCTTT 540  
QY 620 TTGCCCCCTGATGAGGTTGATGACTTACGAGACGCAATTCGTACTTTGGGACTTT 674  
Db 541 TTGCCCCCTGATGAGGTTGATGACTTACGAGACGCAATTCGTACTTTGGGACTTT 595

RESULT 28  
AF529841  
LOCUS  
DEFINITION Cricetulus griseus clone 1 transgenic Homo sapiens AID (AID) mRNA,  
complete cds.  
AF529841 597 bp mRNA linear SYN 19-AUG-2002

```
ACCESSION AF529841
VERSION AF529841.1 GI:22297269
SOURCE Chinese hamster.
ORGANISM Cricetulus griseus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
Cricetulus.
REFERENCE 1 (bases 1 to 597)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 597)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
    source
        1..597
            /organism="Cricetulus griseus"
            /db_xref="taxon:10029"
            /clone="1"
            /cell_line="CHO"
            /transgenic
        1..597
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
        1..597
            /gene="AID"
        1..597
            /gene="AID"
            /codon_start=1
            /product="AID"
            /protein_id="AAM95428.1"
            /db_xref="GI:22297270"
            /translation="MDSLMMNRKFLYQFNVRWAKGRRETYLCYVVKRRDSATFSFL
            DFGYLRNKGCHVELLFURYISDWLDPGRCYRVTWFTSWSPCYDKARHVAFLRGNP
            NLSRIPTARLYFCEDRKAEPGLRLHRAGVQIAIMTKDYFCWNTFVENHRTFFK
            AWEGLHENSVRSLRRLILLPLYEVDLDRDAFTWGR"
    source
        129 a 164 c 157 g 147 t
BASE COUNT
ORIGIN
Query Match 19.0%; Score 535; DB 12; Length 597;
Best Local Similarity 99.8%; Pred. No. 1.6e-267;
Matches 585; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 80 ATGGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAATGTCGCTGG 139
Db 1 ATGGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAATGTCGCTGG 60
QY 140 GCTAAGGGTCGGCGTGAGACCTACCTGTGCTACGTAGTGAAGAGCGTGACAGTGCTACA 199
Db 61 GCTAAGGGTCGGCGTGAGACCTACCTGTGCTACGTAGTGAAGAGCGTGACAGTGCTACA 120
QY 200 TCCTTTTACTGACCTTGGTTATCTTCGCAATAGAAGCGGTGCCACGTGGAATGGTC 259
Db 121 TCCTTTTACTGACCTTGGTTATCTTCGCAATAGAAGCGGTGCCACGTGGAATGGTC 180
QY 260 TTCCCTCCGTACATCTCGGACTGGGACCTAGACCTTGGCCGCTGCTACCGCGTCACCTGG 319
Db 181 TTCCCTCCGTACATCTCGGACTGGGACCTAGACCTTGGCCGCTGCTACCGCGTCACCTGG 240
QY 320 TTCACCTCTCGAGCCCTGCTACGACTGTGCCGACATGTGCCGACTTCTTGGGAGG 379
Db 241 TTCACCTCTCGAGCCCTGCTACGACTGTGCCGACATGTGCCGACTTCTTGGGAGG 300
QY 380 AACCCCAACCTCAGTGTGAGGATCTTCACCGCGCCCTCTACTTCTGTGAGGACCGCAAG 439
Db 301 AACCCCAACCTCAGTGTGAGGATCTTCACCGCGCCCTCTACTTCTGTGAGGACCGCAAG 360
QY 440 GCTGAGCCCGAGGGGCTGGCGGGCTGACACCGCGGGGGTGCATAATAGCCATCATGACC 499
|||||
```

```
Db 361 GCTGAGCCCGAGGGGCTGCGCGGCTGCACCGCGGGTGCAACATAGTCATCATGAC 420
QY 500 TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAAACCATGAAGAATTTCAA 559
|||||
Db 421 ATCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAAACCATGAAGAATTTCAA 480
QY 560 GCCTGGGAAGGCGCTGCATGAAAATTCAGTTCGCTCTCTCCAGACAGCTTCGGTGCAT 619
|||||
Db 481 GCCTGGGAAGGCGCTGCATGAAAATTCAGTTCGCTCTCTCCAGACAGCTTCGGTGCAT 540
QY 620 TTGCCCCCTATGAGGTTGATGACATTACGAGACGCAATTTTCGTACTT 665
|||||
Db 541 TTGCCCCCTATGAGGTTGATGACATTACGAGACGCAATTTTCGTACTT 586
RESULT 29
AF529843
LOCUS Cricetulus griseus clone 3 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION complete cds.
ACCESSION AF529843
VERSION AF529843.1 GI:22297273
KEYWORDS Chinese hamster.
SOURCE Cricetulus griseus
ORGANISM Cricetulus griseus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
Cricetulus.
REFERENCE 1 (bases 1 to 597)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 597)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
    source
        1..597
            /organism="Cricetulus griseus"
            /db_xref="taxon:10029"
            /clone="3"
            /cell_line="CHO"
            /transgenic
        1..597
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
        1..597
            /gene="AID"
        1..597
            /gene="AID"
            /codon_start=1
            /product="AID"
            /protein_id="AAM95430.1"
            /db_xref="GI:22297274"
            /translation="MDSLMMNRKFLYQFNVRWAKGRRETYLCYVVKRRDSATFSFL
            DFGYLRNKGCHVELLFURYISDWLDPGRCYRVTWFTSWSPCYDKARHVAFLRGNP
            NLSRIPTARLYFCEDRKAEPGLRLHRAGVQIAIMTKDYFCWNTFVENHRTFFK
            AWEGLHENSVRSLRRLILLPLYEVDLDRDAFTWGR"
    source
        129 a 164 c 156 g 148 t
BASE COUNT
ORIGIN
Query Match 19.0%; Score 535; DB 12; Length 597;
Best Local Similarity 99.8%; Pred. No. 1.6e-267;
Matches 585; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 80 ATGGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAATGTCGCTGG 139
Db 1 ATGGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAATGTCGCTGG 60
QY 140 GCTAAGGGTCGGCGTGAGACCTACCTGTGCTACGTAGTGAAGAGCGTGACAGTGCTACA 199
|||||
```

Db 61 GCTAAGGCTCGCGGTGAGACCTACCTGTGCTACCTAGTGAAGAGCGGTGACAGTCTTACA 120  
Qy 200 TCCTTTTTCACCTGGACTTTGGTTATCTTCGCAATAGAACGCTGCCACGTGGAATTGCTC 259  
Db 121 TCCTTTTTCACCTGGACTTTGGTTATCTTCGCAATAGAACGCTGCCACGTGGAATTGCTC 180  
Qy 260 TTCTTCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCCGCTGCTACCGCGTCACTGG 319  
Db 181 TTCTTCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCCGCTGCTACCGCGTCACTGG 240  
Qy 320 TTCACCTCTCGAGCCCTCTAGACATGTGCGCGACATGTGGCGGACTTTCTTGGGAGG 379  
Db 241 TTCACCTCTCGAGCCCTCTAGACATGTGCGCGACATGTGGCGGACTTTCTTGGGAGG 300  
Qy 380 AACCCCAACCTCAGTCTCAGGATCTTACCGCGGCTGACCTGTGAGGACCGCAAG 439  
Db 301 AACCCCAACCTCAGTCTCAGGATCTTACCGCGGCTGACCTGTGAGGACCGCAAG 360  
Qy 440 GCTGAGCCGAGGGCTCGCGGCTGCACCGCGGCTGCAAAATAGCCATCATGACC 499  
Db 361 GCTGAGCCGAGGGCTCGCGGCTGCACCGCGGCTGCAAAATAGCCATCATGACC 420  
Qy 500 TTCAAGATTTATTTTACTGCTGGAATCTTTTGTAGAAAACCATGAAAGAACTTTCAA 559  
Db 421 TTCAAGATTTATTTTACTGCTGGAATCTTTTGTAGAAAACCATGAAAGAACTTTCAA 480  
Qy 560 GCCTGGGAGGGCTGCATGAAATTCAGTTCGTCTCTCCAGACAGCTTCGGCGCATCCTT 619  
Db 481 GCCTGGGAGGGCTGCATGAAATTCAGTTCGTCTCTCCAGACAGCTTCGGCGCATCCTT 540  
Qy 620 TTGCCCCGTATGAGGTTGATGACATTACGAGACGCAATTCGTACTT 665  
Db 541 TTGCCCCGTATGAGGTTGATGACATTACGAGACGCAATTCGTACTT 586

RESULT 30  
AF529844  
LOCUS  
DEFINITION  
Cricetulus griseus clone 4 transgenic Homo sapiens AID (AID) mRNA,  
complete cds.  
ACCESSION  
AF529844  
VERSION  
AF529844.1 GI:22297275  
KEYWORDS  
Chinese hamster.  
SOURCE  
Cricetulus griseus  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
Cricetulus.  
REFERENCE  
1 (bases 1 to 597)  
Martin,A. and Scharff,M.D.  
AUTHORS  
Somatic hypermutation of the AID transgene in B cells and non-B  
cells  
JOURNAL  
Unpublished  
REFERENCE  
2 (bases 1 to 597)  
Martin,A. and Scharff,M.D.  
AUTHORS  
Direct Submission  
TITLE  
Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of  
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA  
JOURNAL  
Location/Qualifiers  
FEATURES  
1..597  
source  
/organism="Cricetulus griseus"  
/db\_xref="taxon:10029"  
/clone="4"  
/cell\_line="CHO"  
/transgenic  
1..597  
source  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
1..597  
gene  
/gene="AID"  
1..597  
CDS  
/gene="AID"  
/codon\_start=1

/product="AID"  
/protein\_id="AA095431.1"  
/db\_xref="GI:22297276"  
/translation="MDSLLNNRRKFLYQFNVRWAKGRRETYLCYVYVVRKRRDSATSFSL  
DFGLYLRNKGCHVELLEFLRYISMDLDPGRYVTFWTSWSPCYDCARHVAADFURGNP  
NLSURIFTLARLYFCEDRKAEPGRRLRHAGVQIAIMTFKDYFYCWNTFEVNHETERK  
AWEGHNSVRLSRQLRILLPLVEVDLRDAFTWGR"  
BASE COUNT 128 a 163 c 158 g 148 t  
ORIGIN  
Query Match 19.0%; Score 535; DB 12; Length 597;  
Best Local Similarity 99.8%; Pred. No. 1.6e-267;  
Matches 585; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 80 ATGCACAGCCCTCTGATGACACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGGCTGG 139  
Db 1 ATGCACAGCCCTCTGATGACACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGGCTGG 60  
Qy 140 GCTAAGGCTCGCGGTGAGACCTACCTGTGTCTACCTAGTGAAGAGCGGTGACAGTCTCTACA 199  
Db 61 GCTAAGGCTCGCGGTGAGACCTACCTGTGTCTACCTAGTGAAGAGCGGTGACAGTCTCTACA 120  
Qy 200 TCCTTTTTCACCTGGACTTTGGTTATCTTCGCAATAGAACGCTGCCACGTGGAATTGCTC 259  
Db 121 TCCTTTTTCACCTGGACTTTGGTTATCTTCGCAATAGAACGCTGCCACGTGGAATTGCTC 180  
Qy 260 TTCTCTCGCTACATCTCGGACTGGGACCTAGACCCCTGCACCGCTGCACCGCTCACCTGG 319  
Db 181 TTCTCTCGCTACATCTCGGACTGGGACCTAGACCCCTGCGCCGCTGCTACCGCGTCACTGG 240  
Qy 320 TTCACCTCTCGGAGCCCTCTGCTACGACTGTGCCCGACATGTGGCGGACTTTCTTGGAGGG 379  
Db 241 TTCACCTCTCGGAGCCCTCTGCTACGACTGTGCCCGACATGTGGCGGACTTTCTTGGAGGG 300  
Qy 380 AACCCCAACCTCAGTCTGAGGATCTTACCGCGGCGCTCTACTTCTGTGAGGACCGCAAG 439  
Db 301 AACCCCAACCTCAGTCTGAGGATCTTACCGCGGCGCTCTACTTCTGTGAGGACCGCAAG 360  
Qy 440 GCTGAGCCGAGGGCTCGCGGCTGCACCGCGGCTGCAAAATAGCCATCATGACC 499  
Db 361 GCTGAGCCGAGGGCTCGCGGCTGCACCGCGGCTGCAAAATAGCCATCATGACC 420  
Qy 500 TTCAAGATTTATTTTACTGCTGGAATCTTTTGTAGAAAACCATGAAAGAACTTTCAA 559  
Db 421 TTCAAGATTTATTTTACTGCTGGAATCTTTTGTAGAAAACCATGAAAGAACTTTCAA 480  
Qy 560 GCCTGGGAGGGCTGCATGAAATTCAGTTCGTCTCTCCAGACAGCTTCGGCGCATCCTT 619  
Db 481 GCCTGGGAGGGCTGCATGAAATTCAGTTCGTCTCTCCAGACAGCTTCGGCGCATCCTT 540  
Qy 620 TTGCCCCGTATGAGGTTGATGACATTACGAGACGCAATTCGTACTT 665  
Db 541 TTGCCCCGTATGAGGTTGATGACATTACGAGACGCAATTCGTACTT 586

RESULT 31  
AF529845  
LOCUS  
DEFINITION  
Cricetulus griseus clone 5 transgenic Homo sapiens AID (AID) mRNA,  
complete cds.  
ACCESSION  
AF529845  
VERSION  
AF529845.1 GI:22297277  
KEYWORDS  
Chinese hamster.  
SOURCE  
Cricetulus griseus  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
Cricetulus.  
REFERENCE  
1 (bases 1 to 597)  
Martin,A. and Scharff,M.D.  
AUTHORS  
Somatic hypermutation of the AID transgene in B cells and non-B  
cells  
JOURNAL  
Unpublished

```
REFERENCE 2 (bases 1 to 597)
AUTHORS Martin.A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
LOCATION/Qualifiers
FEATURES
source
1..597
/organism="Cricetulus griseus"
/db_xref="taxon:10029"
/clone="5"
/cell_line="CHO"
/transgenic
1..597
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..597
/gene="AID"
1..597
/gene="AID"
/protein_id="AAM95432.1"
/product="AID"
/db_xref="GI:22297278"
/translation="MDSLMMNRKFLYQFNVRWAKGRRETYLCVYVVKRRDSATFSFL
DFGYLRNKGCHVELLFURYISDWLDPGRCRYRVTWFTSWSPCYDCARHVADFLRGNP
NLSLRIFTARLYFCEDRKAEPGLRRLHRAGVOIAIMTFKDYFCWNTFVENHIERTEK
AWEGIHENSVRSLRQLRLLPLLYEVDDLDRDAFTWGR"
BASE COUNT 128 a 163 c 157 g 149 t
ORIGIN
Query Match 19.0%; Score 535; DB 12; Length 597;
Best Local Similarity 99.8%; Pred. No. 1.6e-267;
Matches 585; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 80 ATGGACAGCCCTTGTATGACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGCTGG 139
Db |||||||
1 ATGGACAGCCCTTGTATGACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGCTGG 60
QY 140 GCTAAGGTCGGCGTGGACCTACCTGCTGCTAGTGAAGAGCGGTGACAGTGTATACA 199
Db |||||||
61 GCTAAGGTCGGCGTGGACCTACCTGCTGCTAGTGAAGAGCGGTGACAGTGTATACA 120
QY 200 TCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGACGGCTGCCAGTGGAAATGTGTC 259
Db |||||||
121 TCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGACGGCTGCCAGTGGAAATGTGTC 180
QY 260 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCGCTGCTACCGGCTCACCTGG 319
Db |||||||
181 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCGCTGCTACCGGCTCACCTGG 240
QY 320 TTCACCTCCTGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACTTTCTGCGAGGG 379
Db |||||||
241 TTCACCTCCTGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACTTTCTGCGAGGG 300
QY 380 AACCCCAACCTCAGTCTCAGGATCTTACCGGGCGGCTCTACTTCTGTGAGGACCGGAAG 439
Db |||||||
301 AACCCCAACCTCAGTCTCAGGATCTTACCGGGCGGCTCTACTTCTGTGAGGACCGGAAG 360
QY 440 GCTGAGCCGAGGGCTCGCGGCTGCACCGCGCGGGGTGCCAATAGCCATCATGACC 499
Db |||||||
361 GCTGAGCCGAGGGCTCGCGGCTGCACCGCGGGGTGCCAATAGCCATCATGACC 420
QY 500 TTCAAGATATTTTACTGCTGGAATACTTTTGTAGAAAACCAATGAAGAACTTTCAAA 559
Db |||||||
421 TTCAAGATATTTTACTGCTGGAATACTTTTGTAGAAAACCAATGAAGAACTTTCAAA 480
QY 560 GCCTGGGAAGGGCTGCATGAATAATTCAGTCTCTCCAGACAGCTTCGGGGCATTCCTT 619
Db |||||||
481 GCCTGGGAAGGGCTGCATGAATAATTCAGTCTCTCCAGACAGCTTCGGGGCATTCCTT 540
QY 620 TTGCCCCCTGTATGAGGTTGATGACTTACGACGCGCATTTCTGACTT 665
Db |||||||
541 TTGCCCCCTGTATGAGGTTGATGACTTACGACGCGCATTTCTGACTT 586
```

```
RESULT 32
LOCUS AF529848
DEFINITION Cricetulus griseus clone 8 transgenic Homo sapiens AID (AID) mRNA,
complete cds.
ACCESSION AF529848
VERSION AF529848.1 GI:22297283
KEYWORDS Chinese hamster.
SOURCE Cricetulus griseus
ORGANISM Cricetulus griseus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
Cricetulus
REFERENCE 1 (bases 1 to 597)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 597)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
LOCATION/Qualifiers
FEATURES
source
1..597
/organism="Cricetulus griseus"
/db_xref="taxon:10029"
/clone="8"
/cell_line="CHO"
/transgenic
1..597
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..597
/gene="AID"
1..597
/gene="AID"
/protein_id="AAM95435.1"
/product="AID"
/codon_start=1
/db_xref="GI:22297284"
/translation="MDSLMMNRKFLYQFNVRWAKGRRETYLCVYVVKRRDSATFSFL
DFGYLRNKGCHVELLFURYISDWLDPGRCRYRVTWFTSWSPCYDCARHVADFLRGNP
NLSLRIFTARLYFCEDRKAEPGLRRLHRAGVOIAIMTFKDYFCWNTFVENHIERTEK
AWEGIHENSVRSLRQLRLLPLLYEVDDLDRDAFTWGR"
BASE COUNT 127 a 164 c 158 g 148 t
ORIGIN
Query Match 19.0%; Score 535; DB 12; Length 597;
Best Local Similarity 99.8%; Pred. No. 1.6e-267;
Matches 585; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 80 ATGGACAGCCCTTGTATGACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGCTGG 139
Db |||||||
1 ATGGACAGCCCTTGTATGACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGCTGG 60
QY 140 GCTAAGGTCGGCGTGGACCTACCTGCTGCTAGTGAAGAGCGGTGACAGTGTATACA 199
Db |||||||
61 GCTAAGGTCGGCGTGGACCTACCTGCTGCTAGTGAAGAGCGGTGACAGTGTATACA 120
QY 200 TCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGACGGCTGCCAGTGGAAATGTGTC 259
Db |||||||
121 TCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGACGGCTGCCAGTGGAAATGTGTC 180
QY 260 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCGCTGCTACCGGCTCACCTGG 319
Db |||||||
181 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCGCTGCTACCGGCTCACCTGG 240
QY 320 TTCACCTCCTGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACTTTCTGCGAGGG 479
Db |||||||
241 TTCACCTCCTGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACTTTCTGCGAGGG 300
```

QY 380 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGGCTCTACTTCTGTGAGGACCGCAAG 439  
|||||  
Db 301 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGGCTCTACTTCTGTGAGGACCGCAAG 360  
QY 440 GCTGAGCCCGAGGGCTCGCGGCTGCGACCGCGCGGGTGCAAAATAGCCATCATGACC 499  
Db 361 GCTGAGCCCGAGGGCTCGCGGCTGCGACCGCGCGGGTGCAAAATAGCCATCATGACC 420  
QY 500 TTCAAGAGATATTTTACTGCTGGAATACATTTTGTAGAAAACCATGAAAGAACTTTCAAA 559  
|||||  
Db 421 TTCAAGAGATATTTTACTGCTGGAATACATTTTGTAGAAAACCATGAAAGAACTTTCAAA 480  
QY 560 GCCTGGGAAGGGTGCGTGAATAATTCAGTTGCTCTCTCCAGACGCTTCGGCGCATCCTT 619  
Db 481 GCCTGGGAAGGGTGCGTGAATAATTCAGTTGCTCTCTCCAGACGCTTCGGCGCATCCTT 540  
QY 620 TTGCCCCCTGTAGAGTTGATGACTTACGAGACGCGCATTTTCGTACTT 665  
Db 541 TTGCCCCCTGTAGAGTTGATGACTTACGAGACGCGCATTTTCGTACTT 586  
RESULT 33  
AF529849  
LOCUS AF529849 597 bp mRNA linear SYN 19-AUG-2002  
DEFINITION Cricetus griseus clone 9 transgenic Homo sapiens AID (AID) mRNA,  
complete cds.  
ACCESSION AF529849  
VERSION AF529849.1 GI:22297285  
KEYWORDS  
SOURCE Chinese hamster.  
ORGANISM Cricetus griseus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
Cricetulus.  
REFERENCE 1 (bases 1 to 597)  
AUTHORS Martin, A. and Scharff, M.D.  
TITLE Somatic hypermutation of the AID transgene in B cells and non-B  
cells  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 597)  
AUTHORS Martin, A. and Scharff, M.D.  
TITLE Direct Submission  
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of  
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA  
FEATURES  
source  
1. .597  
/organism="Cricetus griseus"  
/db\_xref="taxon:10029"  
/clone="9"  
/cell\_line="CHO"  
/transgenic  
source  
1. .597  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
gene  
1. .597  
/gene="AID"  
CDS  
1. .597  
/gene="AID"  
/codon\_start=1  
/product="AID"  
/protein\_id="AA095436.1"  
/db\_xref="GI:22297286"  
/translation="MDSLNNRRKFLYQFNKRWAKRRRETYLYVVKRRDSATSFSL  
DFGLNKGCHVELFLRYISDWLDPGRYRVWTWSWPCYDCARHVDPLRGNP  
NLSRLFTARLYFCGRKAEPEGLRLHRAGVQIAIMTEKDYFYCHNTFVENHRTFK  
ANEGLHNSVRLSRQLRLRLPLLYEVDLDLDAFTWGR"  
BASE COUNT 127 a 164 c 158 g 148 t  
ORIGIN  
Query Match 19.0%; Score 535; DB 12; Length 597;  
Best Local Similarity 99.8%; Pred. No. 1.6e-267;  
Matches 585; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 80 ATGCACAGCCTCTTGATGAACCGGAGAGTCTTTTACCAATTCAAAATGTCCGCTGG 139  
|||||  
Db 1 ATGCACAGCCTCTTGATGAACCGGAGAGTCTTTTACCAATTCAAAATGTCCGCTGG 60  
QY 140 GCTAAGGGTGGCGTGAGACCTACCTGTGCTAGTGAAGAGGGGTGACAGTGTCTACA 199  
Db 61 GCTAAGGGTGGCGTGAGACCTACCTGTGCTAGTGAAGAGGGGTGACAGTGTCTACA 120  
QY 200 TCCCTTTTCACTGACATTTGGTTATCTTCCAAATGAAGACGGCTGCCACGTGGAAATTCCTC 259  
|||||  
Db 121 TCCCTTTTCACTGACATTTGGTTATCTTCCAAATGAAGACGGCTGCCACGTGGAAATTCCTC 180  
QY 260 TTTCTCTCCGTACATCTCGGACTGGGACTAGACCCCTGGCGCTGCTACCCGCTCACTCGG 319  
|||||  
Db 181 TTTCTCTCCGTACATCTCGGACTGGGACTAGACCCCTGGCGCTGCTACCCGCTCACTCGG 240  
QY 320 TTTACCTCTCGAGCCCTGCTACGACTGTGCCGACATGTGGCCGACTTTTCGCGAGGG 379  
Db 241 TTTACCTCTCGAGCCCTGCTACGACTGTGCCGACATGTGGCCGACTTTTCGCGAGGG 300  
QY 380 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGGCTCTACTTCTGTGAGGACCGCAAG 439  
|||||  
Db 301 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGGCTCTACTTCTGTGAGGACCGCAAG 360  
QY 440 GCTGAGCCCGAGGGTGCGCGGCTGCGACCGCGCGGGTGCAAAATAGCCATCATGACC 499  
Db 361 GCTGAGCCCGAGGGTGCGCGGCTGCGACCGCGCGGGTGCAAAATAGCCATCATGACC 420  
QY 500 TTCAAGAGATATTTTACTGCTGGAATACATTTTGTAGAAAACCATGAAAGAACTTTCAAA 559  
Db 421 TTCAAGAGATATTTTACTGCTGGAATACATTTTGTAGAAAACCATGAAAGAACTTTCAAA 480  
QY 560 GCCTGGGAAGGGTGCGTGAATAATTCAGTTGCTCTCTCCAGACGCTTCGGCGCATCCTT 619  
Db 481 GCCTGGGAAGGGTGCGTGAATAATTCAGTTGCTCTCTCCAGACGCTTCGGCGCATCCTT 540  
QY 620 TTGCCCCCTGTAGAGTTGATGACTTACGAGACGCGCATTTTCGTACTT 665  
Db 541 TTGCCCCCTGTAGAGTTGATGACTTACGAGACGCGCATTTTCGTACTT 586  
RESULT 34  
AF529850  
LOCUS AF529850 597 bp mRNA linear SYN 19-AUG-2002  
DEFINITION Cricetus griseus clone 10 transgenic Homo sapiens AID (AID) mRNA,  
complete cds.  
ACCESSION AF529850  
VERSION AF529850.1 GI:22297287  
KEYWORDS  
SOURCE Chinese hamster.  
ORGANISM Cricetus griseus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
Cricetulus.  
REFERENCE 1 (bases 1 to 597)  
AUTHORS Martin, A. and Scharff, M.D.  
TITLE Somatic hypermutation of the AID transgene in B cells and non-B  
cells  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 597)  
AUTHORS Martin, A. and Scharff, M.D.  
TITLE Direct Submission  
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of  
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA  
FEATURES  
source  
1. .597  
/organism="Cricetus griseus"  
/db\_xref="taxon:10029"  
/clone="10"  
/cell\_line="CHO"  
/transgenic  
1. .597  
source

gene	/organism="Homo sapiens"									
	/db_xref="taxon:9606"									
	1..597									
	/gene="AID"									
CDS	1..597									
	/gene="AID"									
	/codon_start=1									
	/product="Aid"									
FEATURES	/protein_id="AAM95437.1"									
	/db_xref="GI:22297288"									
	/translation="MDSLMMNRKKFLYQKNVRWAKGRRETYLCYVVKRRDSATSFSL DFCYLRNKGCHVELLFLRYISDWLDPPVCYRVTWFTSWSPCYDCARHVADFLKGNP NLSLRIFLARLYFCEDRKAEPGLRLHRAGVOIAIMTFKDYFCWNTFVENHERTEFK ANEGLHENSVRLSRLLPLYLEVDDLRDAFRTWGR"									
	BASE COUNT	128	a	164	c	156	g	149	t	
ORIGIN										
Query Match										
Best Local Similarity 99.8%; Score 535; DB 12; Length 597;										
Matches 585; Conservative 0; Mismatches 1; Indels 0; Gaps 0;										
QY	80	ATGGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGCCTGG	139							
Db	1	ATGGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGCCTGG	60							
QY	140	GCTAAGGGTCGGCGTGAGACCTACCTGTGCTAGCTAGTGAAGAGGCTGACAGTGCTACA	199							
Db	61	GCTAAGGGTCGGCGTGAGACCTACCTGTGCTAGTGAAGAGGCTGACAGTGCTACA	120							
QY	200	TCCTTTTACTGACATTTGGTTATCTTCGCAATAAGAACGGTGCCACGTGGAATTCGTC	259							
Db	121	TCCTTTTACTGACATTTGGTTATCTTCGCAATAAGAACGGTGCCACGTGGAATTCGTC	180							
QY	260	TTCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCGGCTGCTACCGCTCACCTGG	319							
Db	181	TTCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGTCGCTGCTACCGGCTCACCTGG	240							
QY	320	TTACCTCTCGAGCCCTGCTACGACTGTGCCGACATGTGCCGACTTTCTGCGAGGG	379							
Db	241	TTACCTCTCGAGCCCTGCTACGACTGTGCCGACATGTGCCGACTTTCTGCGAGGG	300							
QY	380	AACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCTCTACTTCTGTGAGGACCGCAAG	439							
Db	301	AACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCTCTACTTCTGTGAGGACCGCAAG	360							
QY	440	GCTGAGCCCGAGGGGTGCGGGGCTGCACCGCGCGGGTGCAAAATAGCCATCATGACC	499							
Db	361	GCTGAGCCCGAGGGGTGCGGGGCTGCACCGCGCGGGTGCAAAATAGCCATCATGACC	420							
QY	500	TTCAAGATATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAAGAACTTTTCAA	559							
Db	421	TTCAAGATATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAAGAACTTTTCAA	480							
QY	560	GCTGGGAAGGGCTGCATGAAAATTCAGTTCGCTCTCCAGACAGCTTCGGCGCATCTT	619							
Db	481	GCTGGGAAGGGCTGCATGAAAATTCAGTTCGCTCTCCAGACAGCTTCGGCGCATCTT	540							
QY	620	TTGCCCTCTATGAGGTTTACGACTTACGAGACGCATTTTCGTACTT	665							
Db	541	TTGCCCTCTATGAGGTTTACGACTTACGAGACGCATTTTCGTACTT	586							
RESULT 35										
AF529851										
LOCUS										
DEFINITION										
Cricetulus griseus clone 11 transgenic Homo sapiens AID (AID) mRNA, complete cds.										
ACCESSION										
AF529851										
VERSION										
AF529851.1										
KEYWORDS										
SOURCE										
Chinese hamster.										
Cricetulus griseus										
ORGANISM										
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;										

REFERENCE	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetidae;									
	Cricetus									
	1 (bases 1 to 597)									
	Martin,A. and Scharff,M.D.									
AUTHORS	Somatic hypermutation of the AID transgene in B cells and non-B									
	cells									
	Unpublished									
	2 (bases 1 to 597)									
JOURNAL	Martin,A. and Scharff,M.D.									
	Direct Submission									
	Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of									
	Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA									
FEATURES	Location/Qualifiers									
	1..597									
	/organism="Cricetus griseus"									
	/db_xref="taxon:10029"									
SOURCE	/clone="11"									
	/cell_line="CHO"									
	/transgenic									
	1..597									
source	/organism="Homo sapiens"									
	/db_xref="taxon:9606"									
	1..597									
	/gene="AID"									
gene	1..597									
	/gene="AID"									
	1..597									
	/codon_start=1									
CDS	/product="AID"									
	/protein_id="AAM95438.1"									
	/db_xref="GI:22297290"									
	/translation="MDSLMMNRKKFLYQKNVRWAKGRRETYLCYVVKRRDSATSFSL									
BASE COUNT	DFCYLRNKGCHVELLFLRYISDWLDPQRCYRVTWFTSWSPCYDCARIVADFLKGNP									
	NLSLRFTARLYFCEDRKAEPSELRLHRAGVQIAIMTFKDYFCWNTVENHERTEFK									
	ANEGLHENSVRPLPROLRLILLPLYEVDLDRDAFRTWGR"									
	128	a	165	c	157	g	147	t		
ORIGIN										
Query Match										
Best Local Similarity 99.8%; Score 535; DB 12; Length 597;										
Matches 585; Conservative 0; Mismatches 1; Indels 0; Gaps 0;										
QY	80	ATGGACAGCCTCTTGATGAACCGGAGGAAAGTTTCTTTACCAATTCAAAAATGTCGCCTGG	139							
Db	1	ATGGACAGCCTCTTGATGAACCGGAGGAAAGTTTCTTTACCAATTCAAAAATGTCGCCTGG	60							
QY	140	GCTAAGGGTCGGCGTGAGACCTACCTGTGCTAGCTAGTGAAGAGGCTGACAGTGCTACA	199							
Db	61	GCTAAGGGTCGGCGTGAGACCTACCTGTGCTAGTGAAGAGGCTGACAGTGCTACA	120							
QY	200	TCCTTTTACTGACATTTGGTTATCTTCGCAATAAGAACGGTGCCACGTGGAATTCGTC	259							
Db	121	TCCTTTTACTGACATTTGGTTATCTTCGCAATAAGAACGGTGCCACGTGGAATTCGTC	180							
QY	260	TTCTCCGCTACATCTCGACGTGGGACCTAGACCTGGCGGCTGCTACCGCTCACTGG	319							
Db	181	TTCTCCGCTACATCTCGACGTGGGACCTAGACCTGGCGGCTGCTACCGCTCACTGG	240							
QY	320	TTACCTCTCGAGCCCTGCTACGACTGTGCCGACATGTGGCCGACTTTCTGCGAGGG	379							
Db	241	TTACCTCTCGAGCCCTGCTACGACTGTGCCGACATGTGGCCGACTTTCTGCGAGGG	300							
QY	380	AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG	439							
Db	301	AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG	360							
QY	440	GCTGACCCGAGGGGTGCGGGGCTGCACCGCGCGGGTGCAATAGCCATCATGACC	499							
Db	361	GCTGACCCGAGGGGTGCGGGGCTGCACCGCGCGGGTGCAATAGCCATCATGACC	420							
QY	500	TTCAAGATATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAAGAACTTTTCAA	559							
Db	421	TTCAAGATATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAAGAACTTTTCAA	480							
QY	560	GCTGGGAAGGGCTGCATAAATTCAGTTCGCTCTCCAGACAGCTTCGGCGCATCTT	619							
Db	481	GCTGGGAAGGGCTGCATAAATTCAGTTCGCTCTCCAGACAGCTTCGGCGCATCTT	540							
QY	620	TTGCCCTCTATGAGGTTTACGACTTACGAGACGCATTTTCGTACTT	665							
Db	541	TTGCCCTCTATGAGGTTTACGACTTACGAGACGCATTTTCGTACTT	586							



```
QY 560 GCCTGGGAAGGCGTCATGAAATTCAGTTCTGCTCTCCAGACAGCTTCGCGCATCCTT 619
|||||
Db 481 GCCTGGGAAGGCGTCATGAAATTCAGTTCTGCTCTCCAGACAGCTTCGCGCATCCTT 540

QY 620 TTGCCCCGTATGAGTTGATGACTTACGAGACGATTTCTGACTT 665
|||||
Db 541 TTGCCCCGTATGAGTTGATGACTTACGAGACGATTTCTGACTT 586

RESULT 36
AF529852 597 bp mRNA linear SYN 19-AUG-2002
LOCUS Cricetulus griseus clone 12 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION complete cds.
ACCESSION AF529852
VERSION AF529852.1 GI:22297291
KEYWORDS Chinese hamster.
SOURCE Cricetulus griseus
ORGANISM Cricetulus griseus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
Cricetulus.
REFERENCE 1 (bases 1 to 597)
AUTHORS Martin, A. and Scharff, M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 597)
AUTHORS Martin, A. and Scharff, M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source
1..597
/organism="Cricetulus griseus"
/db_xref="taxon:10029"
/clone="12"
/cell_line="CHO"
/transgenic
1..597
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..597
/gene="AID"
1..597
/gene="AID"
/codon_start=1
/product="AID"
/protein_id="AAM95439.1"
/db_xref="GI:22297292"
/translation="MDSLMMNRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSL
DFGLRNKNGCHVELLFRLYISDWDLPGRCYRVTWFTSWSPCDCAHRAVDFLRGNP
NLSLRIFTARLYFCEDRKAEPGLRLHRAQVQIAIMTFKDYFCWNTFVENHRTFK
AWEGHENSVRLSRQLRLLPLYEVDLDRDAFTWGR"
BASE COUNT 129 a 164 c 156 g 148 t
ORIGIN

Query Match 19.0%; Score 535; DB 12; Length 597;
Best Local Similarity 99.8%; Pred. No. 1.6e-267;
Matches 585; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 80 ATGGACACGCTCTTGATGAACCGGAGGAAGTTTCTTTACCAATTCAAAATGTCGCGTGG 139
|||||
Db 1 ATGGACACGCTCTTGATGAACCGGAGGAAGTTTCTTTACCAATTCAAAATGTCGCGTGG 60

QY 140 GCTAAGGTCGCGGTGAGACCTACTGCTGCTAGCTAGTGAAGCGCTGACAGTGTCTACA 199
|||||
Db 61 GCTAAGGTCGCGGTGAGACCTACTGCTGCTAGCTAGTGAAGCGCTGACAGTGTCTACA 120

QY 200 TCCCTTTTCACGTGGACTTTGGTTATCTTCGCAATAAGAACGGCTGCACGCTGGAATTGCTC 259
|||||
Db 121 TCCCTTTTCACGTGGACTTTGGTTATCTTCGCAATAAGAACGGCTGCCACGCTGGAATTGCTC 180
```

```
QY 260 TTCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGSCCGCTGCTACCGCGTCACCTGG 319
|||||
Db 181 TTCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGSCCGCTGCTACCGCGTCACCTGG 240

QY 320 TTCACCTCTCGGAGCCCTGCTAGACTGTGCCGACATGTGGCGGACTTTCTTGGGAGG 379
|||||
Db 241 TTCACCTCTCGGAGCCCTGCTAGACTGTGCCGACATGTGGCGGACTTTCTTGGGAGG 300

QY 380 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 439
|||||
Db 301 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360

QY 440 GCTGAGCCCGAGGCGCTGCGCGCTGCACCGCGCGGGGTGCAAAATAGCCATCATGACC 499
|||||
Db 361 GCTGAGCCCGAGGACTGCGCGCTGCACCGCGCGGGGTGCAAAATAGCCATCATGACC 420

QY 500 TTCAAAGATTATTTTACTGCTGGAATCTTTCTAGAAAACCATGAAGAATTTTCAA 559
|||||
Db 421 TTCAAAGATTATTTTACTGCTGGAATCTTTCTAGAAAACCATGAAGAATTTTCAA 480

QY 560 GCCTGGGAAGGCGTCATGAAATTCAGTTCTCTCCAGACAGCTTCGCGCATCCTT 619
|||||
Db 481 GCCTGGGAAGGCGTCATGAAATTCAGTTCTCTCCAGACAGCTTCGCGCATCCTT 540

QY 620 TTGCCCTGTATGAGTTGATGACTTACGAGACGATTTCTGACTT 665
|||||
Db 541 TTGCCCTGTATGAGTTGATGACTTACGAGACGATTTCTGACTT 586

RESULT 37
AF529853 597 bp mRNA linear SYN 19-AUG-2002
LOCUS Cricetulus griseus clone 13 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION complete cds.
ACCESSION AF529853
VERSION AF529853.1 GI:22297293
KEYWORDS Chinese hamster.
SOURCE Cricetulus griseus
ORGANISM Cricetulus griseus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
Cricetulus.
REFERENCE 1 (bases 1 to 597)
AUTHORS Martin, A. and Scharff, M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 597)
AUTHORS Martin, A. and Scharff, M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source
1..597
/organism="Cricetulus griseus"
/db_xref="taxon:10029"
/clone="13"
/cell_line="CHO"
/transgenic
1..597
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..597
/gene="AID"
1..597
/gene="AID"
/codon_start=1
/product="AID"
/protein_id="AAM95440.1"
/db_xref="GI:22297294"
/translation="MDSLMMNRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSL
DFGLRNKNGCHVELLFRLYISDWDLPGRCYRVTWFTSWSPCDCAHRAVDFLRGNP
NLSLRIFTARLYFCEDRKAEPGLRLHRAQVQIAIMTFKDYFYWNTFVENHRTFK
```

BASE COUNT		129 a	164 c	156 g	148 t
ORIGIN		AWEGLHNSVRLSRQLRILLPLYEVDLDRDAFRTWGR"			
Query Match		19.0%;	Score 535;	DB 12;	Length 597;
Best Local Similarity		99.8%;	Pred. No. 1.6e-267;		
Matches 585;		Conservative	0;	Mismatches	1; Indels 0; Gaps 0;
Qy	80	ATGGACAGCCTCTTCATGAACCGGAGGAAGTTTCTTTACCAATTCAAAAATGTCGCGTGG	139		
Db	1	ATGGACAGCCTCTTCATGAACCGGAGGAAGTTTCTTTACCAATTCAAAAATGTCGCGTGG	60		
Qy	140	GCTAAGGTCGCGTGAGACCTACCTGTGCTAGTGTAGAGAGCGGTGACAGTGTCTACA	199		
Db	61	GCTAAGGTCGCGTGAGACCTACCTGTGCTAGTGTAGAGAGCGGTGACAGTGTCTACA	120		
Qy	200	TCCTTTTTCACGTGGAGTGTGGTGTATCTTCGCAATAAGAACGGGTGCCAGTGTGAATGTCTC	259		
Db	121	TCCTTTTTCACGTGGAGTGTGGTGTATCTTCGCAATAAGAACGGGTGCCAGTGTGAATGTCTC	180		
Qy	260	TTCTCTCCGCTACATCTCGGACTTGGGACCTAGACCTGTGCGCGGCGCTCTACTTCTGTGAGACCGCAAG	319		
Db	181	TTCTCTCCGCTACATCTCGGACTTGGGACCTAGACCTGTGCGCGGCGCTCTACTTCTGTGAGACCGCAAG	240		
Qy	320	TTGACCTCTCGAGAGCCCTGTGACACTGTGCGCGACATGTGGCGGACTTTTGTGAGAGG	379		
Db	241	TTGACCTCTCGAGAGCCCTGTGACACTGTGCGCGACATGTGGCGGACTTTTGTGAGAGG	300		
Qy	380	AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGAGACCGCAAG	439		
Db	301	AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGAGACCGCAAG	160		
Qy	440	GCTGAGCCGAGGGCTGCGGCGGTGCACCGCGCGGTGCAAAATAGCAATATGACCC	499		
Db	361	GCTGAGCCGAGGGCTGCGGCGGTGCACCGCGCGGTGCAAAATAGCAATATGACCC	420		
Qy	500	TTCAAAGATTATTTTACTGCTGGAATACATTTTGTAGAAAACCATGAAAGACTTTTCAA	559		
Db	421	TTCAAAGATTATTTTACTGCTGGAATACATTTTGTAGAAAACCATGAAAGACTTTTCAA	480		
Qy	560	GCCTGGGAAGGGCTGCATGAAATTCAGTTCGTCTCCAGACAGCTTCCGACTT	619		
Db	481	GCCTGGGAAGGGCTGCATGAAATTCAGTTCGTCTCCAGACAGCTTCCGACTT	540		
Qy	620	TTGCCCTGTATGAGGTTGATGACTTACGAGACGCAATTCGCTACTT	665		
Db	541	TTGCCCTGTATGAGGTTGATGACTTACGAGACGCAATTCGCTACTT	586		
RESULT 38					
AF529854					
LOCUS		597 bp	mRNA	linear	SYN 19-AUG-2002
DEFINITION		Cricetulus griseus clone 14 transgenic Homo sapiens AID (AID) mRNA, complete cds.			
ACCESSION		AF529854			
VERSION		AF529854.1 GI:22297295			
KEYWORDS		Chinese hamster.			
SOURCE		Cricetulus griseus			
ORGANISM		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetulus.			
REFERENCE		1 (bases 1 to 597)			
AUTHORS		Martin,A. and Scharff,M.D.			
TITLE		Somatic hypermutation of the AID transgene in B cells and non-B cells			
JOURNAL		Unpublished			
REFERENCE		2 (bases 1 to 597)			
AUTHORS		Martin,A. and Scharff,M.D.			
TITLE		Direct Submission			
JOURNAL		Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA			
FEATURES		Location/Qualifiers			
BASE COUNT		128 a	163 c	158 g	148 t
ORIGIN					
Query Match		19.0%;	Score 535;	DB 12;	Length 597;
Best Local Similarity		99.8%;	Pred. No. 1.6e-267;		
Matches 585;		Conservative	0;	Mismatches	1; Indels 0; Gaps 0;
Qy	80	ATGGACAGCCTCTTCATGAACCGGAGGAAGTTTCTTTACCAATTCAAAAATGTCGCGTGG	139		
Db	1	ATGGACAGCCTCTTCATGAACCGGAGGAAGTTTCTTTACCAATTCAAAAATGTCGCGTGG	60		
Qy	140	GCTAAGGTCGCGTGAGACCTACCTGTGCTAGTGTAGAGAGCGGTGACAGTGTCTACA	199		
Db	61	GCTAAGGTCGCGTGAGACCTACCTGTGCTAGTGTAGAGAGCGGTGACAGTGTCTACA	120		
Qy	200	TCCTTTTTCACGTGGAGTGTGGTGTATCTTCGCAATAAGAACGGGTGCCAGTGTGAATGTCTC	259		
Db	121	TCCTTTTTCACGTGGAGTGTGGTGTATCTTCGCAATAAGAACGGGTGCCAGTGTGAATGTCTC	180		
Qy	260	TTCTCTCCGCTACATCTCGGACTTGGGACCTAGACCTGTGCGCGGCGCTCTACTTCTGTGAGACCGCAAG	319		
Db	181	TTCTCTCCGCTACATCTCGGACTTGGGACCTAGACCTGTGCGCGGCGCTCTACTTCTGTGAGACCGCAAG	240		
Qy	320	TTCACTCTCTGGAGCCCTGTGACACTGTGCGCGACATGTGCCGAGCTTTTGTGAGAGG	379		
Db	241	TTCACTCTCTGGAGCCCTGTGACACTGTGCGCGACATGTGCCGAGCTTTTGTGAGAGG	300		
Qy	380	AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGAGACCGCAAG	439		
Db	301	AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGAGACCGCAAG	360		
Qy	440	GCTGAGCCGAGGGCTGCGGCGGTGCACCGCGCGGTGCAAAATAGCCATCATGACC	499		
Db	361	GCTGAGCCGAGGGCTGCGGCGGTGCACCGCGCGGTGCAAAATAGCCATCATGACC	420		
Qy	500	TTCAAGATTATTTTACTGCTGGAATACATTTTGTAGAAAACCATGAAAGACTTTTCAA	559		
Db	421	TTCAAGATTATTTTACTGCTGGAATACATTTTGTAGAAAACCATGAAAGACTTTTCAA	480		
Qy	560	GCCTGGGAAGGGCTGCATGAAATTCAGTTCGTCTCTCCAGACAGCTTCCGCGCATCCTT	619		
Db	481	GCCTGGGAAGGGCTGCATGAAATTCAGTTCGTCTCTCCAGACAGCTTCCGCGCATCCTT	540		
Qy	620	TTGCCCTGTATGAGGTTGATGACTTACGAGACGCAATTCGCTACTT	665		
Db	541	TTGCCCTGTATGAGGTTGATGACTTACGAGACGCAATTCGCTACTT	586		
RESULT 39					
AF529855					
LOCUS		591 bp	mRNA	linear	SYN 19-AUG-2002
DEFINITION		Cricetulus griseus clone 15 transgenic Homo sapiens AID (AID) mRNA.			

```
complete cds.
AF529855.1 GI:22297297
VERSION
KEYWORDS
SOURCE
ORGANISM
Chinese hamster.
Cricetulus griseus
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
Cricetulus.
REFERENCE
1 (bases 1 to 591)
Martin, A. and Scharff, M.D.
Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL
REFERENCE
2 (bases 1 to 591)
Martin, A. and Scharff, M.D.
Direct Submission
TITLE
JOURNAL
Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
LOCATION/Qualifiers
1. .591
/organism="Cricetulus griseus"
/db_xref="taxon:10029"
/clone="15"
/cell_line="CHO"
/transgenic
1. .591
/organism="Homo sapiens"
/db_xref="taxon:9606"
<1..591
/gene="AID"
<1..591
/gene="AID"
/codon_start=1
/product="AID"
/protein_id="AAM95442.1"
/db_xref="GI:22297298"
/translocation="SLMNRRKFLYQKFNVRWAKRRETYLCYVVKRRDSATSFSLDF
GYLRNKGCHVELLFYISDWLDPCGYRVTWFTSWSPCYDCARHVAFLRGNPL
SLRFTARLYFCEDRAEPEGLRHRAGVOIAIMTFKDYFCWNTFVENHRTFAW
EGLHNSVRLSRLRLTLLPLYEVDLDRDAFTLGL"
BASE COUNT 126 a 162 c 155 g 148 t
ORIGIN
Query Match 18.8%; Score 529; DB 12; Length 591;
Best Local Similarity 99.8%; Pred. No. 2.2e-264;
Matches 579; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 86 AGCCTCTTGATGAACCGGAGGAGTTTCTTTACCAATTCAAAATGTCGCTGGGCTAAG 145
Db 1 AGCCTCTTGATGAACCGGAGGAGTTTCTTTACCAATTCAAAATGTCGCTGGGCTAAG 60
Qy 146 GTCGCGGTGAGACCTACCTGTGCTAGTAGTGAAGAGCGGTGACAGTGTACATCCTTT 205
Db 61 GTCGCGGTGAGACCTACCTGTGCTAGTAGTGAAGAGCGGTGACAGTGTACATCCTTT 120
Qy 206 TCACTGGACTTTGGTTATCTTCGCAATAAAGACGGCTGCCACGTGGGAATGCTTCCCTC 265
Db 121 TCACTGGACTTTGGTTATCTTCGCAATAAAGACGGCTGCCACGTGGGAATGCTTCCCTC 180
Qy 266 CGCTACATCTCGGACTGGGACCTAGACCCCTGGCGGCTGCTACCGGTCACCTGGTTCAAC 325
Db 181 CGCTACATCTCGGACTGGGACCTAGACCCCTGGCGGCTGCTACCGGTCACCTGGTTCAAC 240
Qy 326 TCCTGGAGCCCTCTACGACTGTGCCGACATGTGGCCGACTTCTTCCGAGGGAACCCC 385
Db 241 TCCTGGAGCCCTCTACGACTGTGCCGACATGTGGCCGACTTCTTCCGAGGGAACCCC 300
Qy 386 AACCTCAGTCTGAGGATCTTCAACCGCGGCTCTTACTTCTGTGAGGACCGCAAGGCTGAG 445
Db 301 AACCTCAGTCTGAGGATCTTCAACCGCGGCTCTTACTTCTGTGAGGACCGCAAGGCTGAG 360
Qy 446 CCCGAGGGGCTGGCGGCTGTGCACCGCGCGGGGTGCAAAATAGCCATCATGACCTTCAAA 505
```

```
|||||
361 CCGAGGGCTGGCGGCTGCACCCGCGGGTGCAATAGCCATCATGACTTCAAA 420
Qy 506 GATTATTTTACTGCTGGAACTACTTTGTAGAAAACCATGAAAGAACTTTTCAAGCCTGG 565
Db 421 GATTATTTTACTGCTGGAACTACTTTGTAGAAAACCATGAAAGAACTTTTCAAGCCTGG 480
Qy 566 GAAGGCTGCATGAAATTCAGTTCGTCTCTCCAGACAGCTTCGGCGCATCTTTTGGCCC 625
Db 481 GAAGGCTGCATGAAATTCAGTTCGTCTCTCCAGACAGCTTCGGCGCATCTTTTGGCCC 540
Qy 626 CTGTATGAGTTGATGACTTACGAGAGCGCATTTTCGTACTT 665
Db 541 CTGTATGAGTTGATGACTTACGAGAGCGCATTTTCGTACTT 580

RESULT 40
AF529840 577 bp mRNA linear SYN 19-AUG-2002
LOCUS
DEFINITION
Mus musculus clone 13 transgenic Homo sapiens AID (AID) mRNA,
complete cds.
ACCESSION
AF529840
VERSION
AF529840.1 GI:22297267
KEYWORDS
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE
1 (bases 1 to 577)
Martin, A. and Scharff, M.D.
Somatic hypermutation of the AID transgene in B and non-B cells
JOURNAL
REFERENCE
2 (bases 1 to 577)
Martin, A. and Scharff, M.D.
Direct Submission
TITLE
JOURNAL
Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
LOCATION/Qualifiers
1. .577
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="13"
/cell_type="hybridoma pl-5"
/transgenic
1. .577
/organism="Homo sapiens"
/db_xref="taxon:9606"
<1..577
/gene="AID"
<1..577
/gene="AID"
/codon_start=3
/product="AID"
/protein_id="AAM95427.1"
/db_xref="GI:22297268"
/translocation="RRKFLYQKFNVRWAKRRETYLCYVVKRRDSATSFSLDFGLRN
KNGCHVELLFYISDWLDPCGYRVTWFTSWSPCYDCARHVAFLRGNPLSLRIF
TARLYFCEDRAEPEGLRHRAGVOIAIMTFKDYFCWNTFVENHRTFKAWEGLHE
NSVRLSRLRLTLLPLYEVDLDRDAFTLGL"
BASE COUNT 122 a 159 c 150 g 146 t
ORIGIN
Query Match 18.7%; Score 526; DB 12; Length 577;
Best Local Similarity 99.8%; Pred. No. 8e-263;
Matches 576; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 99 ACCGGAGGAAGTTTCTTTACCAATTCAAAATGTCGCTGGCTAAGGTCGGCGTGAGA 158
Db 1 ACCGGAGGAAGTTTCTTTACCAATTCAAAATGTCGCTGGCTAAGGTCGGCGTGAGA 60
Qy 159 CCTACCTGCTACGTAGTGAAGAGCGGTGACAGTGTACATCTTTTCACTGACCTTG 218
Db 61 CCTACCTGCTACGTAGTGAAGAGCGGTGACAGTGTACATCTTTTCACTGACCTTG 120
```



gene /clone="Ramos 3"  
1..>596  
/gene="AID"  
CDS 1..>596  
/gene="AID"  
/note="Integrated into Burkitt's lymphoma cell line Ramos"  
/codon\_start=1  
/product="AID"  
/db\_xref="GI:22297222"  
/translation="MDSLMMNRKFLYOFKNVRWAKGPRETYLCYVVKRRDSATSFSL  
DFGLRNKNGCHVELLFERYISDWDLDPGRCYRVWTFWSPCDCAHVAIDLGNP  
NLSURIFARLYFCEDRKAEPGRURLRAGVQIAIMTFKDDFYCWNTFVENHERTFK  
AWEGHENSRLRLLPLVEVDLDAFRTLGL"  
BASE COUNT 128 a 164 c 156 g 148 t

Query Match 17.5%; Score 494; DB 9; Length 596;  
Best Local Similarity 99.7%; Pred. No. 3.9e-246;  
Matches 594; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 80 ATGGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGCGTGG 139  
DB 1 ATGGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGCGTGG 60  
QY 140 GCTAAGGTCGCGGTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGTCTACA 199  
DB 61 GCTAAGGTCGCGGTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGTCTACA 120  
QY 200 TCCTTTTTCACGTGGACTTTGGTTATCTTCGCAATAGAAGCGGTGCCAGTGAATTGCTC 259  
DB 121 TCCTTTTTCACGTGGACTTTGGTTATCTTCGCAATAGAAGCGGTGCCAGTGAATTGCTC 180  
QY 260 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCTGCGCGCTGTACCGGTACCTGG 319  
DB 181 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCTGCGCGCTGTACCGGTACCTGG 240  
QY 320 TTCACCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGGCGGACTTCTCGGAGGG 379  
DB 241 TTCACCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGGCGGACTTCTCGGAGGG 300  
QY 380 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGACCGCAAG 439  
DB 301 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGACCGCAAG 360  
QY 440 GCTGAGCCCGAGGGGCTGCGGCGCTGCACCGCGCGGTGCAATAGCCATCATGACC 499  
DB 361 GCTGAGCCCGAGGGGCTGCGGCGCTGCACCGCGCGGTGCAATAGCCATCATGACC 420  
QY 500 TTCAAAGATATTATTTTACTGCTGGAATACTTTGTAGAAACCATGAAGAAGTTCAAA 559  
DB 421 TTCAAAGATATTATTTTACTGCTGGAATACTTTGTAGAAACCATGAAGAAGTTCAAA 480  
QY 620 TTGCCCTCTATGAGGTTGATGACTTACGAGACGCAATTCGTACTTTGGGACTTTG 675  
DB 541 TTGCCCTCTATGAGGTTGATGACTTACGAGACGCAATTCGTACTTTGGGACTTTG 596

RESULT 43  
AF529832  
LOCUS  
DEFINITION AF529832 596 bp mRNA linear SVN 19-AUG-2002  
Mus musculus clone 5 transgenic Homo sapiens AID (AID) mRNA,  
complete cds.  
ACCESSION AF529832  
VERSION AF529832.1 GI:22297251  
KEYWORDS  
SOURCE house mouse.  
ORGANISM Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 596)  
Martin, A. and Scharff, M.D.  
Somatic hypermutation of the AID transgene in B and non-B cells  
Unpublished  
2 (bases 1 to 596)  
Martin, A. and Scharff, M.D.  
Direct Submission  
Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of  
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA  
FEATURES  
Location/Qualifiers  
1..596  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="5"  
/cell\_type="hybridoma PI-5"  
/transgenic  
1..596  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
1..>596  
/gene="AID"  
1..>596  
/gene="AID"  
/codon\_start=1  
/product="AID"  
/protein\_id="AAM95419.1"  
/db\_xref="GI:22297252"  
/translation="MDSLMMNRKFLYOFKNVRWAKGPRETYLCYVVKRRDSATSFSL  
DFGLRNKNGCHVELLFERYISDWDLDPGRCYRVWTFWSPCDCAHVAIDLGNP  
NLSURIFARLYFCEDRKAEPGRURLRAGVQIAIMTFKDDFYCWNTFVENHERTFK  
AWEGHENSRLRLLPLVEVDLDAFRTLGL"  
BASE COUNT 129 a 164 c 153 g 150 t  
ORIGIN  
Query Match 17.5%; Score 494; DB 12; Length 596;  
Best Local Similarity 99.7%; Pred. No. 3.9e-246;  
Matches 594; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 80 ATGGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGCGTGG 139  
DB 1 ATGGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGCGTGG 60  
QY 140 GCTAAGGTCGCGGTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGTCTACA 199  
DB 61 GCTAAGGTCGCGGTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGTCTACA 120  
QY 200 TCCTTTTTCACGTGGACTTTGGTTATCTTCGCAATAGAAGCGGTGCCAGTGAATTGCTC 259  
DB 121 TCCTTTTTCACGTGGACTTTGGTTATCTTCGCAATAGAAGCGGTGCCAGTGAATTGCTC 180  
QY 260 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCTGCGCGCTGTACCGGTACCTGG 319  
DB 181 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCTGCGCGCTGTACCGGTACCTGG 240  
QY 320 TTCACCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGGCGGACTTCTCGGAGGG 379  
DB 241 TTCACCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGGCGGACTTCTCGGAGGG 300  
QY 380 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGACCGCAAG 439  
DB 301 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGACCGCAAG 360  
QY 440 GCTGAGCCCGAGGGGCTGCGGCGGTGCACCGCGCGGTGCAATAGCCATCATGACC 499  
DB 361 GCTGAGCCCGAGGGGCTGCGGCGGTGCACCGCGCGGTGCAATAGCCATCATGACC 420  
QY 500 TTCAAAGATATTATTTTACTGCTGGAATACTTTGTAGAAACCATGAAGAAGTTCAAA 559  
DB 421 TTCAAAGATATTATTTTACTGCTGGAATACTTTGTAGAAACCATGAAGAAGTTCAAA 480  
QY 560 GCCTGGGAGGGGCTGATGAAAATTTCAGTTCTCTCCACAGACGCTTCGGGCGATCCTT 619  
|||||

```
Db      481 GCCTGGGAAGGCTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGGCACATCCCT 540
Qy      620 TTGCCCTCTATGAGCTTGATGACTTACGAGACGCAATTCGTACTTTGGGACTTTG 675
      |||||||
Db      541 TTGCCCTCTATGAGCTTGATGACTTACGAGACGCAATTCGTACTTTGGGACTTTG 596

RESULT 44
AF529836      596 bp      mRNA      linear      SYN 19-AUG-2002
LOCUS      Mus musculus clone 9 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION      complete cds.
ACCESSION      AF529836
VERSION      AF529836.1 GI:22297259
KEYWORDS      house mouse.
SOURCE      Mus musculus
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 596)
AUTHORS      Martin,A. and Scharff,M.D.
TITLE      Somatic hypermutation of the AID transgene in B and non-B cells
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 596)
AUTHORS      Martin,A. and Scharff,M.D.
TITLE      Direct Submission
JOURNAL      Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
      Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES      Location/Qualifiers
      source      1..596
      /organism="Mus musculus"
      /db_xref="taxon:10090"
      /clone="9"
      /cell_type="hybridoma P1-5"
      /transgenic
      source      1..596
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      gene      1..>596
      /gene="AID"
      CDS      1..>596
      /gene="AID"
      /codon_start=1
      /product="AID"
      /protein_id="AAM95423.1"
      /db_xref="GI:22297260"
      /translation="MDSLMMNRKFLYQFKNVRWSKGRRETYLCYVVKRRDSATSFSL
      DGYLRNNGCHVELLFRLYISDWLDPCRCRYRVHTFTSWSPCYDCARHVADFLRGNP
      NLSLRFTARLYFCEDRKAEPGLRLRHAGVOIAIMTKDYFYCWNTFVENHERTEFK
      AWEGLHENSVRLSRLRLPLLYEVDLDRDAFTLGL"
      BASE COUNT      127 a 164 c 153 g 152 t
      ORIGIN
      Query Match      17.5%; Score 494; DB 12; Length 596;
      Best Local Similarity      99.7%; Pred. No. 3.9e-246;
      Matches 594; Conservative      0; Mismatches      2; Indels      0; Gaps      0;

Qy      80 ATGNACACCTCTTGATGAACCGGAGAGAGTTCCTTTACCAATTCAAAATGTCGCTGG 139
      |||||||
Db      1 ATGNACACCTCTTGATGAACCGGAGAGAGTTCCTTTACCAATTCAAAATGTCGCTGG 60

Qy      140 GCTAAGGTCGGCTGAGACCTACCTGTGCTACGTAGTGAAGAGCGCTGACAGTGCTACA 199
      |||||||
Db      61 TCTAAGGTCGGCTGAGACCTACCTGTGCTACGTAGTGAAGAGCGCTGACAGTGCTACA 120

Qy      200 TCCCTTTTACGTGGACTTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGAAATGCTC 259
      |||||||
Db      121 TCCCTTTTACGTGGACTTTGGTTATCTTCGCAATAATAACGGCTGCCACGTGGAAATGCTC 180

Qy      260 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCGCTGCTACCGCGTCAACCTGG 319
      |||||||
Db      181 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCGCTGCTACCGCGTCAACCTGG 240
```

```
Qy      320 TTACCTCTCTGAGAGCCCTGCTAGACTGTGCGCAGACATGTGGCGGACTTTCTGCGAGGG 379
      |||||||
Db      241 TTACCTCTCTGAGAGCCCTGCTAGACTGTGCGCAGACATGTGGCGGACTTTCTGCGAGGG 300
      |||||||
Qy      380 AACCCCAACCTCAGTCTCAGGATCTTACCCGCGCGCTCTACTTCTGTGAGGACCGCAAG 439
      |||||||
Db      301 AACCCCAACCTCAGTCTCAGGATCTTACCCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360
      |||||||
Qy      440 GCTGAGCCCGAGGGGCTCGCGGCTGCACCGCGCGGGTGCAGAAATAGCCATCATGACC 499
      |||||||
Db      361 GCTGAGCCCGAGGGGCTCGCGGCTGCACCGCGCGGGTGCAGAAATAGCCATCATGACC 420
      |||||||
Qy      500 TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAAGAACTTTCAA 559
      |||||||
Db      421 TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAAGAACTTTCAA 480
      |||||||
Qy      560 GCCTGGGAAGCGCTGCATGAAAATTCAGTTCTCTCCAGACAGCTTCGGCGCATCCTT 619
      |||||||
Db      481 GCCTGGGAAGCGCTGCATGAAAATTCAGTTCTCTCCAGACAGCTTCGGCGCATCCTT 540
      |||||||
Qy      620 TTGCCCTCTATGAGTGTGATGACTTACGAGACGCATTTCTGACTTTGGGACTTTG 675
      |||||||
Db      541 TTGCCCTCTATGAGTGTGATGACTTACGAGACGCATTTCTGACTTTGGGACTTTG 596

RESULT 45
AF529838      596 bp      mRNA      linear      SYN 19-AUG-2002
LOCUS      Mus musculus clone 11 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION      complete cds.
ACCESSION      AF529838
VERSION      AF529838.1 GI:22297263
KEYWORDS      house mouse.
SOURCE      Mus musculus
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 596)
AUTHORS      Martin,A. and Scharff,M.D.
TITLE      Somatic hypermutation of the AID transgene in B and non-B cells
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 596)
AUTHORS      Martin,A. and Scharff,M.D.
TITLE      Direct Submission
JOURNAL      Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
      Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES      Location/Qualifiers
      source      1..596
      /organism="Mus musculus"
      /db_xref="taxon:10090"
      /clone="11"
      /cell_type="hybridoma P1-5"
      /transgenic
      source      1..596
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      gene      1..>596
      /gene="AID"
      CDS      1..>596
      /gene="AID"
      /codon_start=1
      /product="AID"
      /protein_id="AAM95425.1"
      /db_xref="GI:22297264"
      /translation="MDSLMMNRKFLYQFKNVRWSKGRRETYLCYVVKRRDSATSFSL
      DGYLRNNGCHVELLFRLYISDWLDPCRCRYRVHTFTSWSPCYDCARHVADFLRGNP
      NLSLRFTARLYFCEDRKAEPGLRLRHAGVOIATMTKDYFYCWNTFVENIERTEFK
      AWEGLHENSVRLSRLRLPLLYEVDLDRDAFTLGL"
      BASE COUNT      128 a 163 c 154 g 151 t
      ORIGIN
      Query Match      17.5%; Score 494; DB 12; Length 596;
      Best Local Similarity      99.7%; Pred. No. 3.9e-246;
```

Matches 594; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 80 ATGGACAGCCTCTTGATGAACCGGAGGAAGTTTCTTTACCAATTCAAAAATGTCGCTGG 139  
|||||  
Db 1 ATGGACAGCCTCTTGATGAACCGGAGGAAGTTTCTTTACCAATTCAAAAATGTCGCTGG 60  
QY 140 GCTAAGGGTGGCGGTGAGACCTACCTGTCTCCCAATGAAGAGCGGTGACAGTGCTACA 199  
|||||  
Db 61 GCTAAGGGTGGCGGTGAGACCTACCTGTCTCCCAATGAAGAGCGGTGACAGTGCTACA 120  
QY 200 TCTCTTTTCACTGGACTTTGGTTATCTTCCCAATGAAGAGCGGTGCCACGTGGAATGTCTC 259  
|||||  
Db 121 TCTCTTTTCACTGGACTTTGGTTATCTTCCCAATGAAGAGCGGTGCCACGTGGAATGTCTC 180  
QY 260 TCTCTCGGTACATCTCGGACTGGGACCTAGACCCCTGGCGGCTGTCTACCGCGTCACTCTGG 319  
|||||  
Db 181 TCTCTCGGTACATCTCGGACTGGGACCTAGACCCCTGGCGGCTGTCTACCGCGTCACTCTGG 240  
QY 320 TTTCACTCTCGGAGCCCTGTCTGACGACTGTGCCGACATGTGGCGGACTTTCTTGGAGGG 379  
|||||  
Db 241 TTTCACTCTCGGAGCCCTGTCTGACGACTGTGCCGACATGTGGCGGACTTTCTTGGAGGG 300  
QY 380 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 439  
|||||  
Db 301 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360  
QY 440 GCTGAGCCGAGGGGTGGCGGCTGCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 499  
|||||  
Db 361 GCTGAGCCGAGGGGTGGCGGCTGCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 420  
QY 500 TTCAAGAGATTTTCTGCTGGAATACTTTTGTAGAAAACCATGAAAGACTTTTCAAA 559  
|||||  
Db 421 TTCAAGAGATTTTCTGCTGGAATACTTTTGTAGAAAACCATGAAAGACTTTTCAAA 480  
QY 560 GCCTGGAAGGGCTGCATGAAATTCAGTTCGTCTCTCCAGACAGCTTCGGCGCATCTT 619  
|||||  
Db 481 GCCTGGAAGGGCTGCATGAAATTCAGTTCGTCTCTCCAGACAGCTTCGGCGCATCTT 540  
QY 620 TTGCCCCCTGTATGAGTTTGATGACTTACGAGAGCGCATTTCTGATTTTGGGACTTTG 675  
|||||  
Db 541 TTGCCCCCTGTATGAGTTTGATGACTTACGAGAGCGCATTTCTGATTTTGGGACTTTG 596

RESULT 46  
AF529856 547 bp mRNA linear SYN 19-AUG-2002  
LOCUS Cricetus griseus clone 16 transgenic Homo sapiens AID (AID) mRNA,  
DEFINITION complete cds.  
ACCESSION AF529856  
VERSION AF529856.1 GI:22297299  
KEYWORDS Chinese hamster.  
SOURCE Cricetus griseus  
ORGANISM Cricetus griseus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
Cricetus.  
REFERENCE 1 (bases 1 to 547)  
AUTHORS Martin,A. and Scharff,M.D.  
TITLE Somatic hypermutation of the AID transgene in B cells and non-B  
cells  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 547)  
AUTHORS Martin,A. and Scharff,M.D.  
TITLE Direct Submission  
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of  
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA  
FEATURES Location/Qualifiers  
source 1..547  
/organism="Cricetus griseus"  
/db\_xref="taxon:10029"  
/clone="16"  
/cell\_line="CHO"  
/transgenic

1..547  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
<1..547  
/gene="AID"  
CDS  
1..547  
/gene="AID"  
/codon\_start=2  
/product="AID"  
/protein\_id="AAM95443.1"  
/db\_xref="GI:22297300"  
/translation="VRWAKRRRETYLCYVVKRRDSATSFSLDFGLYLRNKGCHVELLF  
LRYISDWLDPRCYRVYTFWTSWPCYCARHVDLFRGNPNLSLRIFTARLYFCEDR  
KAPEGLRLRHRAVGQIAIMTFKDYFCWNTFVENHRTFKAEWGLHENSRLSRQLR  
RIILPLYEVDDLRLDAFTWGR"  
BASE COUNT 112 a 153 c 147 g 135 t  
ORIGIN  
Query Match 17.3%; Score 487; DB 12; Length 547;  
Best Local Similarity 100.0%; Pred. No. 1.7e-242;  
Matches 487; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 130 TGTCCGCTGGGCTAAGGCTGCGCGTGCAGACCTACCTGTGCTACGTAGTCAAGAGCGTGA 189  
|||||  
Db 1 TGTCCGCTGGGCTAAGGCTGCGCGTGCAGACCTACCTGTGCTACGTAGTCAAGAGCGTGA 60  
QY 190 CAGTGTACATCCTTTTCTCCTGACTTTTGGTTATCTTGCATAAAGAACGGCTGCCACGT 249  
|||||  
Db 61 CAGTGTACATCCTTTTCTCCTGACTTTTGGTTATCTTGCATAAAGAACGGCTGCCACGT 120  
QY 250 GGAATTTGCTTCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCGGCTGTCTACCG 309  
|||||  
Db 121 GGAATTTGCTTCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCGGCTGTCTACCG 180  
QY 310 CGTCACTGCTTCACTCTCTGAGCCCTGCTACGACTGTGCCGACATGTGCCGACTTT 369  
|||||  
Db 181 CGTCACTGCTTCACTCTCTGAGCCCTGCTACGACTGTGCCGACATGTGCCGACTTT 240  
QY 370 TCTGGAGGAAACCCCACTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGA 429  
|||||  
Db 241 TCTGGAGGAAACCCCACTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGA 300  
QY 430 GGACCGAAGGCTGACCCGAGGGGCTGGCGGCTGCACCGCGCGGCTGCAATAGC 489  
|||||  
Db 301 GGACCGAAGGCTGACCCGAGGGGCTGGCGGCTGCACCGCGCGGCTGCAATAGC 360  
QY 490 CATCATGACTTCAAGATTATTTTACTGCTGGAATCTTTTGTAGAAAACCATCAAG 549  
|||||  
Db 361 CATCATGACTTCAAGATTATTTTACTGCTGGAATCTTTTGTAGAAAACCATCAAG 420  
QY 550 AACTTTCAAAGCTGGGAGGCTGCATGAAATTCAGTTCTCTCTCCAGACAGCTTCG 609  
|||||  
Db 421 AACTTTCAAAGCTGGGAGGCTGCATGAAATTCAGTTCTCTCTCCAGACAGCTTCG 480  
QY 610 GCGCATC 616  
|||||  
Db 481 GCGCATC 487  
RESULT 47  
AC007240/c  
LOCUS Homo sapiens BAC clone RP11-83M8 from 2, complete sequence.  
DEFINITION AC007240  
ACCESSION AC007240.2 GI:5306303  
VERSION AC007240.2  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 203250)  
AUTHORS Sulston,J.E. and Waterston,R.  
TITLE Toward a complete human genome sequence





```
/rpt_family="Alu"
repeat_region 17799..18095
/rpt_family="Alu"
repeat_region 18146..18449
/rpt_family="L1"
repeat_region 18450..18745
/rpt_family="Alu"
repeat_region 18746..18928
/rpt_family="L1"
repeat_region 18929..19230
/rpt_family="Alu"
repeat_region 19231..19570
/rpt_family="L1"
repeat_region 19571..19880
/rpt_family="Alu"
repeat_region 19881..20109
/rpt_family="L1"
repeat_region 20110..20415
/rpt_family="Alu"
repeat_region 20416..20446
/rpt_family="L1"
repeat_region 20446..20577
/rpt_family="L1"
repeat_region 20583..20574
/rpt_family="Alu"
repeat_region 21002..21107
/rpt_family="L1"
repeat_region 21464..21698
/rpt_family="L1"
repeat_region 21825..22150
/rpt_family="Alu"
repeat_region 22322..22626
/rpt_family="Alu"
repeat_region 23770..23972
/rpt_family="MER1_type"
repeat_region 24272..24294
/rpt_family="CAAT)n"
repeat_region 24304..24576
/rpt_family="Alu"
repeat_region 24646..24962
/rpt_family="Alu"
repeat_region 24973..25410
```

Query Match 3.1%; Score 86; DB 9; Length 203250;

Best Local Similarity 100.0%; Pred. No. 2.le-33;

Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1970 GCGGTGGTACGAGGACCTGTAATCCAGCTACTTGGAGGCTGAGCGAGGAGTGCCT 2029
|||||
DB 18606 GCGGTGGTACGAGGACCTGTAATCCAGCTACTTGGAGGCTGAGCGAGGAGTGCCT 18547
|||||
QY 2030 TCAACCCAGGAGGTGGAGGTTGCAGT 2055
|||||
DB 18546 TGAACCCAGGAGGTGGAGGTTGCAGT 18521
|||||
```

```
RESULT 48
AC079329 156540 bp DNA linear PRI 20-APR-2002
LOCUS Homo sapiens chromosome 11, clone RP11-265D17, complete sequence.
DEFINITION AC079329
ACCESSION AC079329
VERSION AC079329.15 GI:20128016
KEYWORDS HTC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 156540)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 11, clone RP11-265D17
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 156540)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
```

Anderson,S., Barna,N., Bastien,V., Bada,F., Boguslavkiy,L.,  
Boukhaltier,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,  
Choepe,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,  
DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,  
FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,  
Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,  
Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRocque,K.,  
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,C., Liu,G.,  
Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,  
McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,  
Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,  
O'Donnell,P., O'Neill,D., Olivari,T.M., Oliver,C., Peterson,K.,  
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,  
Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,  
Sognez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo.A.,  
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,  
Zimmer,A. and Zody,M.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Submitted (27-AUG-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 156540)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,  
Boukhaltier,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,  
Chazaro,B., Choepe,Y., Colangelo,M., Collins,S., Collymore,A.,  
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,  
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Jones,C.,  
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,  
Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,  
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,  
McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,  
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,  
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,  
Olivier,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,  
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,  
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,  
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,  
Viel,R., Vo.A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Submitted (10-APR-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
4 (bases 1 to 156540)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,  
Boukhaltier,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,  
Chazaro,B., Choepe,Y., Colangelo,M., Collins,S., Collymore,A.,  
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,  
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,  
Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,  
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,  
McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,  
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,  
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,  
Olivier,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,  
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,  
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,  
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,  
Viel,R., Vo.A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE  
JOURNAL

Submitted (20-APR-2002) Whitehead Institute/MIT Center for Genome

COMMENT      Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Apr 10, 2002 this sequence version replaced gi:19683125.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: <http://www-seq.wi.mit.edu>  
 Contact: [sequence.submissions@genome.wi.mit.edu](mailto:sequence.submissions@genome.wi.mit.edu)  
 ----- Project Information  
 Center project name: L10912  
 Center clone name: 265\_D\_17  
 -----

## FEATURES

source

Location/Qualifiers

1..156540  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="11"  
 /map="11"  
 /clone="RP11-265D17"  
 /clone\_lib="RPC1-11 Human Male BAC"  
 134..139  
 /note="<30 qual SNGL region"  
 813..879  
 /note="single clone coverage"  
 1951..2104  
 /rpt\_family="MIR"  
 complement(2247..2542)  
 /rpt\_family="AluX"  
 complement(2778..2877)  
 /rpt\_family="L3"  
 complement(2885..3564)  
 /rpt\_family="MER104A"  
 complement(3566..4179)  
 /rpt\_family="L3"  
 complement(4218..4323)  
 /rpt\_family="L3"  
 complement(5098..5386)  
 /rpt\_family="AluSg"  
 complement(6406..6481)  
 /rpt\_family="L2"  
 6539..6559  
 /rpt\_family="AT\_Rich"  
 complement(7143..7430)  
 /rpt\_family="L2"  
 7708..7854  
 /rpt\_family="MIR"  
 complement(8134..8319)  
 /rpt\_family="MIR"  
 8584..8784  
 /rpt\_family="MIR"  
 8798..8826  
 /rpt\_family="AT\_Rich"  
 complement(9140..9272)  
 /rpt\_family="MIR"  
 9455..9489  
 /rpt\_family="(T)n"  
 complement(9549..9844)  
 /rpt\_family="AluJo"  
 10846..11158  
 /rpt\_family="AluYb8"  
 11570..11783  
 /rpt\_family="MIR"  
 complement(11746..11929)  
 /rpt\_family="MIR"  
 complement(12908..13215)  
 /rpt\_family="AluX"  
 13878..14048  
 /rpt\_family="MIR"  
 complement(14481..14549)  
 /rpt\_family="MT1K"  
 14554..14673

repeat\_region      /rpt\_family="AluJo"  
 complement(15075..15285)  
 /rpt\_family="MIR"  
 repeat\_region      complement(15300..15549)  
 /rpt\_family="MIR"  
 repeat\_region      complement(15793..15857)  
 /rpt\_family="L2"  
 16085..16320  
 /rpt\_family="MIR"  
 repeat\_region      complement(16384..16463)  
 /rpt\_family="MIR3"  
 17431..17457  
 /rpt\_family="(TG)n"  
 18440..18652  
 /rpt\_family="L2"  
 complement(18894..19027)  
 /rpt\_family="MIR3"  
 repeat\_region      complement(20058..20252)  
 /rpt\_family="MIR3"  
 20274..20323  
 /rpt\_family="L2"  
 repeat\_region      21013..21198  
 /rpt\_family="MIR3"  
 repeat\_region      complement(21757..21948)  
 /rpt\_family="MIR3"  
 repeat\_region      complement(22497..22689)  
 /rpt\_family="MIR"  
 repeat\_region      complement(22767..23110)  
 /rpt\_family="L2"  
 23765..23795  
 /rpt\_family="(GA)n"  
 25844..27056  
 /rpt\_family="L1PA5"  
 repeat\_region      complement(27157..27230)  
 /rpt\_family="L2"  
 repeat\_region      27536..27612  
 /rpt\_family="L2"  
 repeat\_region      complement(27626..27783)  
 /rpt\_family="MERSB"

Query Match      2.8%; Score 79; DB 9; Length 156540;

Best Local Similarity      100.0%; Pred. No. 9.2e-30;

Matches      79; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

Qy      1981    AGGCACCTGTAATCCAGCTACTTGGAGGCTGAGGAGAGAGATCGCTTCAAGCCAGGA    2040  
 Db      32001    AGGCACCTGTAATCCAGCTACTTGGAGGCTGAGGAGAGAGATCGCTTCAAGCCAGGA    32060  
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy      2041    GGTGGAGGTTGCAGTAAGC    2059  
 Db      32061    GGTGGAGGTTGCAGTAAGC    32079  
 |||||||||||||||||||

## RESULT 49

AC023634/c

LOCUS      AC023634      187836 bp      DNA      linear      HTG 16-MAR-2001

DEFINITION      Homo sapiens chromosome 11 clone RP11-133112 map 11, WORKING DRAFT

SEQUENCE, 8 unordered pieces.

AC023634

AC023634.3    GI:13357436

HTG; HTGS\_PHASE1; HTGS\_DRAFT.

Homo sapiens.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE

1 (bases 1 to 187836)

AUTHORS

Birren, B., Linton, L., Nusbaum, C. and Lander, E.

Homo sapiens chromosome 11, clone RP11-133112

JOURNAL

Unpublished

2 (bases 1 to 187836)

REFERENCE

AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,

Anderson, S., Baldwin, J., Barna, N., Bada, F., Boeckmann, J.,

Boukhalil, B., Brown, A., Burkett, G., Campopiano, A., Castiglione, A.,

Choepl,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M., Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Coyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,K., Jones,C., Kann,L., Karatas,A., Klein,J., Landers,T., Largocque,K., Lehoczy,J., Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Menga,V., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliviar,T.M., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,K., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Triglio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission  
Submitted (16-FEB-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Mar 16, 2001 this sequence version replaced gi:7767803.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WTBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence-submissions@genome.wi.mit.edu

----- Project Information

Center project name: L5395

Center clone name: 133.I.12

----- Summary Statistics

Sequencing vector: M13; M77815; 56% of reads

Sequencing vector: Plasmid; n/a; 44% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 179213 bases at least Q40

Consensus quality: 181746 bases at least Q30

Consensus quality: 184176 bases at least Q20

Insert size: 174000; agarose-fp

Insert size: 187136; sum-of-contigs

Quality coverage: 13.0 in Q20 bases; agarose-fp

Quality coverage: 12.1 in Q20.

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 8 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 33786: contig of 33786 bp in length

\* 33787 33886: gap of 100 bp

\* 33887 35099: contig of 1213 bp in length

\* 35100 35199: gap of 100 bp

\* 35200 36514: contig of 1315 bp in length

\* 36515 36614: gap of 100 bp

\* 36615 37990: contig of 1376 bp in length

\* 37991 38090: gap of 100 bp

\* 38091 39599: contig of 1509 bp in length

\* 39600 39699: gap of 100 bp

\* 39700 41135: contig of 1436 bp in length

\* 41136 41235: gap of 100 bp

\* 41236 43376: contig of 2141 bp in length

\* 43377 43476: gap of 100 bp

\* 43477 187836: contig of 144360 bp in length.

Location/Qualifiers

1. .187836

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="11"

FEATURES  
source

/map="11"  
/clone="RP11-133112"  
/clone\_lib="RP11-11 Human Male BAC"  
1. .33786  
/note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:left  
33887. .35099  
/note="assembly\_fragment"  
35200. .36514  
/note="assembly\_fragment"  
36615. .37990  
/note="assembly\_fragment"  
38091. .39599  
/note="assembly\_fragment"  
39700. .41135  
/note="assembly\_fragment"  
41236. .43376  
/note="assembly\_fragment"  
43477. .187836  
/note="assembly\_fragment"  
clone\_end:T7  
vector\_side:right  
BASE COUNT 51537 a 44861 c 42761 g 47977 t 700 others  
ORIGIN

Query Match 2.8%; Score 79; DB 2; Length 187836;  
Best Local Similarity 100.0%; Pred No. 9.2e-30;  
Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1981 AGGCACCTGTAATCCAGCTACTTTGGAGGCTGAGCAGGAGTAATCGCTTGAACCCAGGA 2040  
|||||  
Db 76398 AGGCACCTGTAATCCAGCTACTTTGGAGGCTGAGCAGGAGTAATCGCTTGAACCCAGGA 76339

QY 2041 GGTGGAGGTTGCAGTAAGC 2059  
|||||

Db 76338 GGTGGAGGTTGCAGTAAGC 76320

RESULT 50  
AC004603 41511 bp DNA linear PRI 23-APR-1998  
LOCUS AC004603  
DEFINITION Homo sapiens chromosome 19, cosmid R33496, complete sequence.  
ACCESSION AC004603  
VERSION AC004603.1 GI:3077822  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 41511)  
AUTHORS Lamerdin,J.E., McCreedy,P.M., Skowronski,E., Adamson,A.W., Burkhardt-Schultz,K., Gordon,L., Kyle,A., Ramirez,M., Stilwagen,S., Phan,H., Velasco,N., Ganes,J., Danganan,L., Poundstone,P., Christensen,M., Georgescu,A., Avila,J., Liu,S., Attix,C., Andreise,T., Frankheim,M., Anico-Keller,G., Coefield,J., Duarte,S., Lucas,S., Bruce,R., Thomas,P., Quan,G., Kronmiller,B., Arellano,A., Montgomery,M., Ow,D., Nolan,M., Trong,S., Kobayashi,A., Olsen,A.O. and Carrano,A.V.  
TITLE Sequence analysis of a 2.5 Mb region in 19ql3.2 containing a clustered CEA/PSG gene family  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 41511)  
AUTHORS Lamerdin,J.E.  
TITLE Direct Submission  
JOURNAL Submitted (23-APR-1998) Joint Genome Institute, Lawrence Livermore National Laboratory, 7000 East Ave., Livermore, CA 94551, USA  
COMMENT Map and sequence oriented from q centromere to telomere. Cosmid R33496 overlaps cosmid F19987 to the right (by 2.447 kb) and cosmid F25173 to the right. Additional chr 19 map and sequence information may be obtained at:  
http://www-bio.llnl.gov/bbrp/genome/genome.html.

FEATURES

Location/Qualifiers

```
source
1. .41511
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/map="19q13.2 from BCKDHA-D19S217"
/clone="R33496"
/cell_line="5HL2-B"
/clone_lib="L119NC03 R chromosome 19-specific cosmid library"
/notes="Cosmid library constructed at LLNL from flow-sorted chromosomes from hybrid 5HL2-B, which carries chromosome 19 as its only human chromosome."
/complement(1. .616)
/rpt_family="L1"
612. .938
/rpt_family="L1"
934. .2630
/rpt_family="L1"
3217. .3631
/rpt_family="MSTB"
complement(5887. .5918)
/rpt_family="AT-rich"
complement(6303. .6343)
/rpt_family="(TAAA)n"
complement(6365. .6850)
/rpt_family="AluJo"
6793. .7179
/rpt_family="MLT1B"
9090. .9194
/rpt_family="LINE2"
complement(9305. .9408)
9878. .10178
/rpt_family="FLAW_A"
complement(10336. .10493)
10676. .10792
/rpt_family="LINE2"
10954. .11356
/rpt_family="MSTB"
11886. .12076
/notes="predicted exon, program: grail2exons_human_1.3, frame: 2, quality: good, score: 65.000"
12353. .12393
/rpt_family="MIR"
12727. .12945
/rpt_family="L1ME"
12969. .13742
/rpt_family="L1MB3"
13743. .14041
/rpt_family="AluY"
14061. .14108
/rpt_family="L1MB4"
14567. .14361
/notes="predicted exon, program: grail2exons_human_1.3, frame: 2, quality: good, score: 62.000"
12353. .12393
/rpt_family="MIR"
12727. .12945
/rpt_family="L1ME"
12969. .13742
/rpt_family="L1MB3"
13743. .14041
/rpt_family="AluY"
14061. .14108
/rpt_family="L1MB4"
14567. .14361
/notes="predicted exon, program: grail2exons_human_1.3, frame: 2, quality: good, score: 62.000"
complement(15609. .16217)
/rpt_family="MER67C"
16221. .16426
/rpt_family="AluJo"
16427. .16457
/rpt_family="(TGAA)n"
complement(16458. .16509)
/rpt_family="MER67C"
16548. .16661
/notes="predicted exon, program: grail2exons_human_1.3, frame: 2, quality: good, score: 72.000"
17689. .17990
/rpt_family="AluSx"
18091. .18371
/rpt_family="HERVL"
18445. .18540
/rpt_family="MLT1B"

repeat_region complement(18583. .19157)
/rpt_family="L1"
repeat_region complement(19158. .19458)
/rpt_family="AluY"
19890. .19924
/rpt_family="(CA)n"
complement(19925. .20055)
/rpt_family="MER4_internal"
20076. .20439
/rpt_family="MER50"
complement(20469. .20524)
/rpt_family="(TA)n"
20541. .20823
/rpt_family="MER4_internal"
complement(20869. .21760)
/rpt_family="L1PA8"
22250. .22542
/rpt_family="AluJo"
22812. .23005
/rpt_family="MER4C"
23594. .23733
/notes="predicted exon, program: grail2exons_human_1.3, frame: 1, quality: excellent, score: 83.000"
23753. .23899
/rpt_family="MIR"
complement(24669. .24849)
/rpt_family="MIR"
24950. .24985
/rpt_family="MLT1D"
24987. .25287
/rpt_family="AluSg"
25291. .25469
/rpt_family="MLT1D"
complement(25608. .25638)
/rpt_family="AluSg"
complement(25856. .26042)
/rpt_family="MER11A"
complement(26223. .26594)
/rpt_family="HERVK"
complement(26739. .26902)
/rpt_family="AluSg"
complement(26911. .27277)
/rpt_family="MER57_internal"
27440. .27732
/rpt_family="AluSp"
complement(27759. .29229)
/rpt_family="MER65_internal"
complement(29238. .29530)
/rpt_family="AluY"
complement(29531. .29623)
/rpt_family="MER65_internal"
29819. .29969
/rpt_family="MER5B"
complement(30951. .30994)
/notes="Exon, pregnancy-specific beta-1-glycoprotein 9-like pseudogene. DPS similarity to M32635) pregnancy-specific beta-1-glycoprotein 5 [Homo sapiens] (residues 500. .512); 67% identity."
pseudo
31507. .31749
/rpt_family="LTR16B"
complement(31877. .32061)
/rpt_family="MIR"
complement(32142. .32243)
/rpt_family="MIR"
32244. .32543
/rpt_family="L1"
32546. .32860
/rpt_family="AluJo"
32862. .32986
/rpt_family="L1"
33029. .35071
/rpt_family="L1MB4"
```

```

exon      complement(35975. .36229)
/notes="Exon, pregnancy-specific beta-1-glycoprotein 5-like
pseudogene. Dps similarity to (M32635) pregnancy-specific
beta-1-glycoprotein 5 [Homo sapiens] (residues 415. .499);
91% identity."
/pseudo
exon      complement(36632. .36910)
/notes="Exon, pregnancy-specific beta-1-glycoprotein 5-like
pseudogene. Dps similarity to (M32635) pregnancy-specific
beta-1-glycoprotein 5 [Homo sapiens] (residues 322. .414);
91% identity."
/pseudo
repeat_region complement(37853. .37911)
/rpt_family="LIMB1"
repeat_region 37924. .38055
/rpt_family="LIMB1"
repeat_region 38056. .38218
/rpt_family="MER46"
repeat_region complement(38485. .38798)
/rpt_family="L1"
exon      complement(38984. .39238)
/notes="Exon, pregnancy-specific beta-1-glycoprotein 5-like

Query Match      2.7%; Score 77; DB 9; Length 41511;
Best Local Similarity 100.0%; Pred. No. 1e-28;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1979 GCAGGCACCTGTATCCCGACTTGGGAGGCTGAGGAGGAGTCCGTTGAACCCAG 2038
|||||
Db 17838 GCAGGCACCTGTATCCCGACTTGGGAGGCTGAGGAGGAGTCCGTTGAACCCAG 17897
|||||

QY 2039 GAGGTGGAGGTTGCAGT 2055
|||||
Db 17898 GAGGTGGAGGTTGCAGT 17914
|||||

RESULT 51
AC084034/c
LOCUS
DEFINITION
AC084034 103287 bp DNA linear HTG 09-MAY-2002
HOMO sapiens chromosome 3 clone RP11-502H22, *** SEQUENCING IN
PROGRESS ***, 30 unordered pieces.
AC084034 15 GI:20335777
AC084034.15
HTG: HTGS_PHASE1.
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 103287)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J.J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Devilla,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earhart,D., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlisson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louiseged,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,

```

```

TITLE
JOURNAL
REFERENCE 2 (bases 1 to 103287)
AUTHORS Worley,K.C.
TITLE
JOURNAL
REFERENCE 3 (bases 1 to 103287)
AUTHORS Worley,K.C.
TITLE
JOURNAL
COMMENT
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HCGG
Center clone name: RP11-502H22
----- Summary Statistics
Sequencing vector: M13;
Chemistry: Dye-primer Bodipy: 37% of reads
Chemistry: Dye-terminator Big Dye: 63% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 117633 bases at least Q40
Consensus quality: 136061 bases at least Q30
Consensus quality: 148159 bases at least Q20
Estimated insert size: 141518; sum-of-contigs estimation
Quality coverage: 2x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 30 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 2194: contig of 2194 bp in length
* 2195 2294: gap of unknown length
* 2295 4434: contig of 2140 bp in length
* 4435 4534: gap of unknown length
* 4535 6710: contig of 2176 bp in length
* 6711 6810: gap of unknown length
* 6811 9027: contig of 2217 bp in length
* 9028 9127: gap of unknown length
* 9128 11310: contig of 2183 bp in length
* 11311 11410: gap of unknown length
* 11411 13930: contig of 2520 bp in length
* 13931 14030: gap of unknown length
* 14031 16531: contig of 2501 bp in length

```









repeat\_region 282. .668 /rpt\_family="MaLR"  
repeat\_region 1575. .1615 /rpt\_family="AT\_rich"  
repeat\_region 1616. .1897 /rpt\_family="Alu"  
repeat\_region 3392. .3775 /rpt\_family="Retroviral"  
misc\_feature 4319. .4868 /note="match to EST AI568539 (NID:g4531913) tn40e08.xl1"  
repeat\_region 5673. .5921 /rpt\_family="MER1\_type"  
misc\_feature 8043. .10028 /note="CpG\_island (%GC=68.4, o/e=0.90, #CpGs=214)"  
repeat\_region 8332. .8364 /rpt\_family="A-rich"  
repeat\_region 9202. .9330 /rpt\_family="CCCCG)n"  
repeat\_region 9950. .9970 /rpt\_family="GC\_rich"  
repeat\_region 10090. .10212 /rpt\_family="L1"  
repeat\_region 10986. .11122 /rpt\_family="(TAGA)n"  
repeat\_region 11243. .11360 /rpt\_family="(CAGA)n"  
repeat\_region 12015. .12240 /rpt\_family="MER4-group"  
repeat\_region 12243. .12853 /rpt\_family="MER4-group"  
repeat\_region 12881. .12991 /rpt\_family="MER94"  
repeat\_region 12998. .13524 /rpt\_family="L1"  
repeat\_region 16196. .16425 /rpt\_family="Retroviral"  
repeat\_region 17256. .17613 /rpt\_family="Retroviral"  
repeat\_region 18007. .18149 /rpt\_family="MER2\_type"  
repeat\_region 18224. .18522 /rpt\_family="MER2\_type"  
repeat\_region 18715. .18832 /rpt\_family="L1"  
repeat\_region 19286. .19373 /rpt\_family="L2"  
repeat\_region 20391. .20481 /rpt\_family="MIR"  
repeat\_region 21808. .22111 /rpt\_family="Alu"  
repeat\_region 22147. .22188 /rpt\_family="AT\_rich"  
repeat\_region 22392. .22555 /rpt\_family="MaLR"  
repeat\_region 23127. .23212 /rpt\_family="MaLR"  
repeat\_region 24894. .25083 /rpt\_family="MER4-group"  
repeat\_region 25147. .25237 /rpt\_family="MER4-group"  
repeat\_region 25269. .25598 /rpt\_family="MER4-group"  
repeat\_region 25616. .26607 /rpt\_family="MER4-group"  
repeat\_region 26945. .27063 /rpt\_family="MER4-group"  
repeat\_region 27210. .27502 /rpt\_family="Alu"  
repeat\_region 27503. .27768 /rpt\_family="MER4-group"  
repeat\_region 28822. .29089 /rpt\_family="MER4-group"  
repeat\_region 29111. .29368

repeat\_region 29377. .29464 /rpt\_family="MER4-group?"  
repeat\_region 29748. .29786 /rpt\_family="AT\_rich"  
repeat\_region 29893. .30466 /rpt\_family="L1"  
repeat\_region 30524. .30638 /rpt\_family="(CATA)n"  
repeat\_region 30886. .30997 /rpt\_family="(CA)n"  
repeat\_region 31015. .31125 /rpt\_family="(CAATA)n"  
repeat\_region 31099. .31230 /rpt\_family="(CACAC)n"  
repeat\_region 32285. .33146 /rpt\_family="L1"  
repeat\_region 33240. .33373 /rpt\_family="MER2\_type"  
repeat\_region 33380. .33559 /rpt\_family="L1"  
misc\_feature 35690. .37077 /note="CpG\_island (%GC=63.0, o/e=0.90, #CpGs=106)"  
repeat\_region 37890. .40107 /rpt\_family="MER2\_type"  
repeat\_region 42804. .43109 /rpt\_family="Alu"  
repeat\_region 44304. .44361 /rpt\_family="MIR"  
  
Query Match 2.7%; Score 77; DB 9; Length 112804;  
Best Local Similarity 100.0%; Pred.No.le-28;  
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1979 GCAGGCACCTGTAATCCAGCTACTTGGAGGCTGAGCAGAGAGATCGCTTGAACCCAG 2038  
Db 101773 GCAGGCACCTGTAATCCAGCTACTTGGAGGCTGAGCAGAGAGATCGCTTGAACCCAG 101714  
  
QY 2039 GAGGTGGAGTTGCAGT 2055  
Db 101713 GAGGTGGAGTTGCAGT 101697  
  
RESULT 55  
LOCUS AY052369 121028 bp DNA linear PRI 16-APR-2002  
DEFINITION Homo sapiens PP2A B56 gamma gene, complete cds, alternative splice products.  
ACCESSION AY052369  
VERSION AY052369.1 GI:16303629  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 121028)  
AUTHORS Muneer,S., Ramalingam,V., Wyatt,R., Schultz,R.A., Minna,J.D. and Kamibayashi,C.  
TITLE Genomic organization and mapping of the gene encoding the PP2A B56gamma regulatory subunit  
JOURNAL Genomics 79 (3), 344-348 (2002)  
MEDLINE 21853305  
PUBMED 11863364  
REFERENCE 2 (bases 1 to 121028)  
AUTHORS Muneer,S., Kamibayashi,C. and Minna,J.D.  
TITLE Direct Submission  
JOURNAL Submitted (24-AUG-2001) Hamon Center for Therapeutic Oncology, University of Texas Southwestern Medical Center, 6000 Harry Hines Blvd, Dallas, TX 75390, USA  
FEATURES  
Location/Qualifiers  
1..121028  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

```
/chromosome="14"
/map="14q32.2"
join(2507..2643,49292..49491,74767..74877,75863..75955,
76038..76168,82840..82899,85603..85711,87110..87163,
94322..94492,99005..99132,102192..102293,105004..105076,
110434..110550,117745..120289)
/product="PP2A B56 gamma 3"
join(2507..2643,49292..49491,74767..74877,75863..75955,
76038..76168,82840..82899,85603..85711,87110..87163,
94322..94492,99005..99132,102192..102293,105004..105076,
117745..120289)
/product="PP2A B56 gamma 2"
join(2507..2643,49292..49491,74767..74877,75863..75955,
76038..76168,82840..82899,85603..85711,87110..87163,
94322..94492,99005..99132,102192..102293,105004..105226)
/product="PP2A B56 gamma 1"
2507..2643
/number=1
join(2580..2643,49292..49491,74767..74877,75863..75955,
76038..76168,82840..82899,85603..85711,87110..87163,
94322..94492,99005..99132,102192..102293,105004..105076,
110434..110550,117745..117876)
/notes="protein phosphatase; splice variant 3"
/codon_start=1
/product="PP2A B56 gamma 3"
/protein_id="AAL14778.1"
/db_xref="GI:16303632"
/translation="MVVDAANSNGFPQPVLLHIRDVPPADQEKLFIOKLQCCVLFD
FVSDPLSLKWKVKRAALSEMVEYITHNRNVITEPIYEVVHFAVNMFTLPPSSN
PTGAFFDPEDEPTLEAAMPHQLQVYEFFLFLESDFQPNIAKKYIDQKVLQLEL
FDSQPRDRFLKTLHRIYKGLRAYIRKQINNIYRIYETEHHNGIAELLEIL
GSIINGFALPLKEEHKIFLLKVLPLHKVLSVYPOLAYCVQVLFKDSLTLEPVV
MALLYWPKTHSPKEVMFLNELEITLDVIEPSEFVKIMEPLFROLAKCVSSPHFOVAE
RALLYWNNYETMSLISDAAKILPIMFPSLYRNSKTHWNKTIHGLIYNALKLFMEMNQ
KLFDCTQOQFAEKLEKLEKMKEREAWKIENTAKANPOYTVYQASMTSIPVAMET
DGPLFEDVQLRKVKDPAHQKDPKRAKDRPLALRKSLEPQDPHFKKALEAHCRADEL
ASQDGR"
join(2580..2643,49292..49491,74767..74877,75863..75955,
76038..76168,82840..82899,85603..85711,87110..87163,
94322..94492,99005..99132,102192..102293,105004..105076,
117745..117876)
/notes="protein phosphatase; splice variant 2"
/codon_start=1
/product="PP2A B56 gamma 2"
/protein_id="AAL14778.1"
/db_xref="GI:16303631"
/translation="MVVDAANSNGFPQPVLLHIRDVPPADQEKLFIOKLQCCVLFD
FVSDPLSLKWKVKRAALSEMVEYITHNRNVITEPIYEVVHFAVNMFTLPPSSN
PTGAFFDPEDEPTLEAAMPHQLQVYEFFLFLESDFQPNIAKKYIDQKVLQLEL
FDSQPRDRFLKTLHRIYKGLRAYIRKQINNIYRIYETEHHNGIAELLEIL
GSIINGFALPLKEEHKIFLLKVLPLHKVLSVYPOLAYCVQVLFKDSLTLEPVV
MALLYWPKTHSPKEVMFLNELEITLDVIEPSEFVKIMEPLFROLAKCVSSPHFOVAE
RALLYWNNYETMSLISDAAKILPIMFPSLYRNSKTHWNKTIHGLIYNALKLFMEMNQ
KLFDCTQOQFAEKLEKLEKMKEREAWKIENTAKANPOYTVYQASMTSIPVAMET
DGPLFEDVQLRKVKDPAHQKDPKRAKDRPLALRKSLEPQDPHFKKALEAHCRADEL
ASQDGR"
join(2580..2643,49292..49491,74767..74877,75863..75955,
76038..76168,82840..82899,85603..85711,87110..87163,
94322..94492,99005..99132,102192..102293,105004..105100)
/notes="protein phosphatase; splice variant 1"
/codon_start=1
/product="PP2A B56 gamma 1"
/protein_id="AAL14777.1"
/db_xref="GI:16303630"
/translation="MVVDAANSNGFPQPVLLHIRDVPPADQEKLFIOKLQCCVLFD
FVSDPLSLKWKVKRAALSEMVEYITHNRNVITEPIYEVVHFAVNMFTLPPSSN
PTGAFFDPEDEPTLEAAMPHQLQVYEFFLFLESDFQPNIAKKYIDQKVLQLEL
FDSQPRDRFLKTLHRIYKGLRAYIRKQINNIYRIYETEHHNGIAELLEIL
GSIINGFALPLKEEHKIFLLKVLPLHKVLSVYPOLAYCVQVLFKDSLTLEPVV
MALLYWPKTHSPKEVMFLNELEITLDVIEPSEFVKIMEPLFROLAKCVSSPHFOVAE
RALLYWNNYETMSLISDAAKILPIMFPSLYRNSKTHWNKTIHGLIYNALKLFMEMNQ
KLFDCTQOQFAEKLEKLEKMKEREAWKIENTAKANPOYTVYQASMTSIPVAMET
46492..49491
/number=2
```

```
exon 74767..74877
/number=3
exon 75863..75955
/number=4
exon 76038..76168
/number=5
exon 82840..82899
/number=6
exon 85603..85711
/number=7
exon 87110..87163
/number=8
exon 94322..94492
/number=9
exon 99005..99132
/number=10
exon 102192..102293
/number=11
exon 105004..105076
/number=12
exon 105077..105226
/number=12A
exon 110434..110550
/number=13
exon 117745..120289
/number=14
BASE COUNT 31520 a 26430 c 27189 g 35887 t 2 others
ORIGIN
Query Match 2.7% Score 77; DB 9; Length 121028;
Best Local Similarity 100.0%; Pred. No. le-28;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1979 CGAGCACCCTGTATATCCACGCTACTTTGGAGGCTCGAGCAGAGAGTAATCGCTTGAACCCAG 2038
|||||
Db 114960 CGAGCACCCTGTATATCCACGCTACTTTGGAGGCTCGAGCAGAGAGTAATCGCTTGAACCCAG 115019
|||||
Qy 2039 GAGGTGGAGGTGGCAT 2055
|||||
Db 115020 GAGGTGGAGGTGGCAT 115036
|||||
RESULT 56
AC079944 126295 bp DNA linear PRI 21-JUN-2002
LOCUS Homo sapiens 3 BAC RP11-435F17 (Roswell Park Cancer Institute Human
DEFINITION BAC Library) complete sequence.
ACCESSION AC079944
VERSION AC079944.17 GI:21535888
KEYWORDS HTG.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 126295)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-ouman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbaria,J., Benton,J., Blmage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,K., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davilla,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Ho,X.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Homsli,F., Howard,S., Huber,J., Hulyk,S., Humu,J.,
```

Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovacic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Lounsgue, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhinney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Moore, S., Morgan, M., Moorish, T., Morris, S., Moser, M., Neal, D., Nelson, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokkenko, S., Oguh, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalob, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Zaylor, S.L., Weinstein, G. and Gibbs, R.

# TITLE

## JOURNAL

REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

On Jun 21, 2002 this sequence version replaced gi:21465331.  
 INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email  
[gc-help@bcm.tmc.edu](mailto:gc-help@bcm.tmc.edu)

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

## ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annot.action.html>.

## FEATURES

source	Location/Qualifiers
1..126295	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="3"
	/clone="RP11-435F17"
	/complement(7..87)
	/rpt_family="L1ME"
	/complement(1249..1492)
	/rpt_family="MIR"
	2769..2793
	/rpt_family="(T)n"
	/complement(2813..2980)
	/rpt_family="AluSx"
	/complement(3472..3761)
	/rpt_family="AluJo"
	3834..3872
	/rpt_family="AT-rich"
	4088..4194
	/rpt_family="MLT1B"
	4197..4494
	/rpt_family="AluY"
	4496..4634
	/rpt_family="GA-rich"
	4637..4703
	/rpt_family="MLT1B"
	4707..4988
	/rpt_family="AluSx"
	5110..5302
	/rpt_family="MLT1B"
	/complement(5818..6121)
	/rpt_family="AluSg"
	6658..6768
	/rpt_family="MIR"
	7071..7101
	/rpt_family="AT-rich"
	/complement(9057..9110)
	/rpt_family="MIR"
	/complement(9111..10226)
	/rpt_family="tigger3b"
	10227..10524
	/rpt_family="AluSg"
	/complement(10525..10588)
	/rpt_family="tigger3b"
	/complement(10589..10749)
	/rpt_family="MIR"
	10793..11077
	/rpt_family="L1MC4"
	11090..11398
	/rpt_family="AluSg"
	11587..11616
	/rpt_family="(GAAA)n"
	12951..13268
	/rpt_family="AluSc"
	13282..13577
	/rpt_family="AluSx"
	13569..13669
	/standard_name="176498"
	/complement(13689..13785)
	/rpt_family="MIR"
	/complement(14091..14393)
	/rpt_family="L1MB7"
	/complement(14650..14704)
	/rpt_family="L2"
	15241..15428
	/rpt_family="MIR"

```

repeat_region 15498..15786
/rpt_family="AluSc"
repeat_region 15787..15813
/rpt_family="TAA)n"
repeat_region complement(15848..16146)
/rpt_family="AluSx"
repeat_region 16268..16323
/rpt_family="AT-rich"
repeat_region complement(16639..16778)
/rpt_family="L2"
repeat_region 17560..17580
/rpt_family="AT-rich"
repeat_region complement(19788..20005)
/rpt_family="L2"
repeat_region complement(20100..20384)
/rpt_family="L2"
repeat_region 21645..21936
/rpt_family="AluSx"
repeat_region 22446..22536
/rpt_family="MER94"
repeat_region complement(22567..22692)
/rpt_family="L2"
repeat_region complement(23033..23341)
/rpt_family="MER58B"
repeat_region 23728..23752

```

Query Match 2.7%: Score 77; DB 9; Length 126295;  
 Best Local Similarity 100.0%; Pred. No. 1e-28;  
 Matches 77: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1979 GCAGCACCTGTATCCAGCTACTTGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAG 2038  
 DB 109614 GCAGCACCTGTATCCAGCTACTTGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAG 109673  
 QY 2039 GAGGTGAGGTGTCAGT 2055  
 DB 109674 GAGGTGAGGTGTCAGT 109690

RESULT 57

AC078928  
 LOCUS AC078928 154968 bp DNA linear HTG 05-SEP-2000  
 DEFINITION Homo sapiens chromosome 12q clone RP11-22G4, WORKING DRAFT  
 AC078928  
 VERSION AC078928.6 GI:9966147  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 Muzny,D.M., Adams,C., Bailey,M., Barberia,J., Blankenburg,K., Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C., Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C., David,R., Delgado,D., Deshazo,D., Ding,Y., Domah-Rashid,N., Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D., Forcum-Tansey,J., Frantz,P., Ganesha,R., Gorrell,J.H., Gorrell,L.L., Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hognes,M., Holloway,C., Hosack,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M., Kelly,S., Kondeljowski,N., Kong,Y., Kovar,C., Leal,B., Li,Z., Lichtarge,O., Liu,W., Logan,O., Lozano,R.J., Lu,J., Luchter,R., Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M., Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S., Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L., Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S., Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sugchang,R., Tabor,P., Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wahbah,M., Watlington,S., Weinstock,G., Weinstock,I.R., Williamson,A., Worley,K., Wren,J., Wrensford,G., Yu,W., Zhou,X., Nelson,D. and Gibbs,R.  
 Direct Submission  
 Unpublished

REFERENCE  
 AUTHORS  
 JOURNAL

2 (bases 1 to 154968)  
 Worley,K.C.  
 Direct Submission  
 Submitted (11-AUG-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 On Sep 4, 2000 this sequence version replaced gi:9929481.

COMMENT

----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: http://www.hgsc.bcm.tmc.edu/  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: HCAS  
 Center clone name: RP11-22G4  
 ----- Summary Statistics  
 Sequencing vector: M13: L08821  
 Chemistry: Dye-Primer Bodipy: 80% of reads  
 Chemistry: Dye-terminator Big Dye: 20% of reads  
 Assembly program: Phrap: version 0.990329  
 Consensus quality: 150687 bases at least Q40  
 Consensus quality: 154338 bases at least Q30  
 Consensus quality: 156644 bases at least Q20  
 Estimated insert size: 152379; sum-of-ctnigs estimation  
 Quality coverage: 0x in Q20 bases; agarose-fp estimation  
 Quality coverage: 5x in Q20 bases; sum-of-ctnigs estimation  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank\_draft\_data.html).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 10 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 51941: contig of 51941 bp in length  
 51942 52041: gap of unknown length  
 52042 85871: contig of 33830 bp in length  
 85872 85971: gap of unknown length  
 85972 99094: contig of 13123 bp in length  
 99095 109291: contig of 10097 bp in length  
 109292 109391: gap of unknown length  
 109392 118498: contig of 9107 bp in length  
 118499 118598: gap of unknown length  
 118599 127576: contig of 8977 bp in length  
 127576 127676: gap of unknown length  
 127676 135234: contig of 7559 bp in length  
 135235 135334: gap of unknown length  
 135335 142541: contig of 7207 bp in length  
 142542 142641: gap of unknown length  
 142642 150582: contig of 7941 bp in length  
 150583 150682: gap of unknown length  
 150683 154968: contig of 4286 bp in length.

FEATURES  
 source

1. 154968  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="12q"  
 /clone="RP11-22G4"

BASE COUNT 48695 a 29406 c 27663 g 48296 t 908 others  
 ORIGIN

Query Match 2.7%: Score 77; DB 2; Length 154968;  
 Best Local Similarity 100.0%; Pred. No. 1e-28;  
 Matches 77: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1979 GCAGCACCTGTATCCAGCTACTTGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAG 2038  
 DB 126077 GCAGCACCTGTATCCAGCTACTTGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAG 126136

TITLE  
 JOURNAL

```
QY      2039 GAGGTGGAGGTGGCAGT 2055
        |||||
Db      126137 GAGGTGGAGGTGGCAGT 126153

AC011022      162958 bp      DNA      linear      HTG 26-MAR-2000
Homo sapiens clone RP11-318C12, WORKING DRAFT SEQUENCE, 11
unordered pieces.
AC011022      4      GI:7329290
HTG: HTGS_PHASE1; HTGS_DRAFT.
Homo sapiens
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 162958)
Birren.B., Linton.L., Nusbaum.C. and Lander.E.
Homo sapiens chromosome, clone RP11-318C12
Unpublished
2 (bases 1 to 162958)
Birren.B., Linton.L., Nusbaum.C., Lander.E., Allen.N., Anderson.M.,
Baldwin.J., Barna.N., Beckerly.R., Boguslavskiy.L., Boukhalter.B.,
Brown.A., Castle.A., Colangelo.M., Collins.S., Collymore.A.,
Cooke.P., DeRellano.K., Dewar.K., Domino.M., Donelan.L., Doyle.M.,
Ferreira.P., FitzHugh.W., Forrest.C., Funke.R., Gage.D.,
Galagan.J., Gardyna.S., Grant.G., Hagos.B., Heaford.A., Horton.L.,
Howland.J.C., Johnson.R., Jones.C., Kann.L., Karatas.A., Klein.J.,
Lehoczky.J., Lieu.C., Locke.K., Macdonald.P., Marquis.N.,
McEwan.P., McGurk.A., McKernan.K., McLaughlin.J., Meldrim.J.,
Morrow.J., Naylor.J., Norman.C.H., O'Connor.T., O'Donnell.P.,
Peterson.K., Pollara.V., Riley.R., Roy.A., Santos.R., Severy.P.,
Stange-Thomann.N., Stojanovic.N., Subramanian.A., Talamas.J.,
Testaye.S., Tirrell.A., Vassiliev.H., Vo.A., Wheeler.J., Wu.X.,
Wymann.D., Ye.W.J., Zimmer.A. and Zody.M.
Direct Submission
Submitted (30-SEP-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 26, 2000 this sequence version replaced gi:6630602.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L1785
Center clone name: 318_C_12
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 154951 bases at least Q40
Consensus quality: 159496 bases at least Q30
Consensus quality: 161045 bases at least Q20
Insert size: 168000; agarose-fp
Insert size: 161958; sum-of-contigs
Quality coverage: 4.9 in Q20 bases; agarose-fp
Quality coverage: 5.1 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1546 1645: contig of 1545 bp in length
* 1546 1645: gap of 100 bp

1646 1688: contig of 43 bp in length
1689 1788: gap of 100 bp
1789 13851: contig of 12063 bp in length
13852 13951: gap of 100 bp
13952 27920: contig of 13969 bp in length
27921 28020: gap of 100 bp
28021 41980: contig of 13960 bp in length
41981 42080: gap of 100 bp
42081 57968: contig of 15888 bp in length
57969 58068: gap of 100 bp
58069 73743: contig of 15675 bp in length
73744 73843: gap of 100 bp
73844 87132: contig of 13289 bp in length
87133 87232: gap of 100 bp
87233 105996: contig of 18764 bp in length
105997 106096: gap of 100 bp
106097 127200: contig of 21104 bp in length
127201 127301: gap of 100 bp
127301 162958: contig of 35658 bp in length.

FEATURES
            Location/Qualifiers
                source
                    1..162958
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone="RP11-318C12"
                        /clone_lib="RPC1-11 Human Male BAC"
                    1..1545
                        /note="assembly_fragment"
                    1646..1688
                        /note="assembly_fragment"
                        clone_end:SP6
                        vector_side:right
                    1789..13851
                        /note="assembly_fragment"
                    13952..27920
                        /note="assembly_fragment"
                    28021..41980
                        /note="assembly_fragment"
                    42081..57968
                        /note="assembly_fragment"
                    58069..73743
                        /note="assembly_fragment"
                    73844..87132
                        /note="assembly_fragment"
                    87233..105996
                        /note="assembly_fragment"
                    106097..127200
                        /note="assembly_fragment"
                    127301..162958
                        /note="assembly_fragment"
                        clone_end:T7
                        vector_side:right

BASE COUNT      50339 a 31972 c 32404 g 47242 t 1001 others
ORIGIN

Query Match      2.7%; Score 77; DB 2: Length 162958;
Best Local Similarity 100.0%; Pred. No.1e-28;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1979 GCAGGCACCTGTAATCCCGACTTGGGAGGCTGAGCAGCAGAGTAATCGCTTGAACCCAG 2038
        |||||
Db      106130 GCAGGCACCTGTAATCCCGACTTGGGAGGCTGAGCAGCAGAGTAATCGCTTGAACCCAG 106189

QY      2039 GAGGTGGAGGTGGCAGT 2055
        |||||
Db      106190 GAGGTGGAGGTGGCAGT 106206

RESULT 59
AC007342/c
LOCUS      AC007342      177777 bp      DNA      linear      PRI 29-SEP-2001
DEFINITION      Homo sapiens chromosome 16 clone RP11-44Fl4, complete sequence.
ACCESSION      AC007342
VERSION      AC007342.5      GI:15808502
```

**KEYWORDS**  
**SOURCE** Homo sapiens.  
**ORGANISM** Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
**REFERENCE**  
**AUTHORS** DOE Joint Genome Institute.  
**TITLE** Sequencing of Human Chromosome 16  
**JOURNAL** Unpublished  
**REFERENCE**  
**AUTHORS** 2 (bases 1 to 177777)  
 Bruce, D., Mundt, M., Doggett, N., Munk, C., Saunders, E., Robinson, D.,  
 Jones, M., Buckingham, J., Chasteen, L., Thompson, S., Goodwin, L.,  
 Bryant, J., Tesmer, J., Meincke, L., Longmire, J., White, S., Tatam, O.,  
 Campbell, C., Fawcett, J., Maltbie, M., Bussod, M., Sutherland, R.,  
 McMurry, K., Han, C. and Deaven, L.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (21-APR-1999) Center for Human Genome Studies, DOE Joint  
 Genome Institute, Los Alamos National Laboratory, MS M888, Los  
 Alamos, NM 87545, USA  
**REFERENCE**  
**AUTHORS** 3 (bases 1 to 177777)  
 DOE Joint Genome Institute.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (29-SEP-2001) Production Sequencing Facility, DOE Joint  
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
**COMMENT** On Sep 29, 2001 this sequence version replaced gi:14589412.  
 Sequence Quality Assessment:  
 This entry has been annotated with sequence quality  
 estimates computed by the Phrap assembly program.  
 All manually edited bases have been reduced to quality zero.  
 Quality levels above 40 are expected to have less than  
 1 error in 10,000 bp.  
 Base-by-base quality values are not generally visible from the  
 GenBank flat file format but are available as part  
 of this entry's ASN.1 file.  
 -----  
 Sequence Quality Assessment:  
 This entry has been annotated with sequence quality  
 estimates computed by the Phrap assembly program.  
 All manually edited bases have been reduced to quality zero.  
 Quality levels above 40 are expected to have less than  
 1 error in 10,000 bp.  
 Base-by-base quality values are not generally visible from the  
 GenBank flat file format but are available as part  
 of this entry's ASN.1 file.  
 -----  
**FEATURES**  
 source  
 1. .177777  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="16"  
 /clone="RP11-44F14"  
 BASE COUNT 53238 a 37201 c 36594 g 50744 t  
 ORIGIN  
 Query Match 2.7%; Score 77; DB 9; Length 177777;  
 Best Local Similarity 100.0%; Pred. No. le-28;  
 Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1979 GCAGGCACCTGTAATCCCGCTACTTGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAG 2038  
 Db 100973 GCAGGCACCTGTAATCCCGCTACTTGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAG 100914  
 QY 2039 GAGGTGGAGTTGCAGT 2055  
 Db 100913 GAGGTGGAGTTGCAGT 100897  
 RESULT 60  
 AL391866/c  
 LOCUS  
 DEFINITION Homo sapiens chromosome 1 clone RP11-315J13, \*\*\* SEQUENCING IN  
 PROGRESS \*\*\*; 4 unordered pieces.  
 ACCESSION AL391866 AC073084

**VERSION** AL391866.8 GI:13561275  
**KEYWORDS** HTG; HTGS\_PHASE1; HTGS\_CANCELLED.  
**SOURCE** human.  
**ORGANISM** Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
**REFERENCE**  
**AUTHORS** 1 (bases 1 to 181183)  
 McIay, K.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
**COMMENT** requests: clonerequest@sanger.ac.uk  
 On or before May 15, 2001 this sequence version replaced  
 gi:8901286, gi:10186833.  
 Draft Sequence Produced by Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 http://genome.wustl.edu/gsc/index.shtml  
 -----  
 Center: Genome Center  
 Center code: SC  
 Web site: http://www.sanger.ac.uk  
 Contact: humquery@sanger.ac.uk  
 -----  
 Project Information  
 Center project name: BA315J13  
 -----  
 Summary Statistics  
 Assembly program: XGAP4; version 4.5  
 Sequencing vector: M13; M7815; 46% of reads  
 Sequencing vector: plasmid; L08752; 53% of reads  
 Chemistry: Dye-terminator Big Dye; 53% of reads  
 Chemistry: Dye-primer-amersham; 46% of reads  
 Consensus quality: 179412 bases at least Q40  
 Consensus quality: 179718 bases at least Q30  
 Consensus quality: 180265 bases at least Q20  
 Insert size: 180883; sum-of-contigs  
 Quality coverage: 7.89x in Q20 bases; agarose-fp  
 Quality coverage: 8.15x in Q20 bases; agarose-fp  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 4 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 23472: contig of 23472 bp in length  
 \* 23473 23572: gap of 100 bp  
 \* 23573 60331: contig of 36759 bp in length  
 \* 60332 60431: gap of 100 bp  
 \* 60432 164641: contig of 104210 bp in length  
 \* 164642 164741: gap of 100 bp  
 \* 164742 181183: contig of 16442 bp in length.  
**FEATURES**  
 Location/Qualifiers  
 1. 181183  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="11"  
 /clone="RP11-315J13"  
 /note="assembly\_fragment:01806"  
 fragment\_chain:1  
 23573 60331  
 /note="assembly\_fragment:04020"  
 fragment\_chain:1  
 60432 164641  
 /note="assembly\_fragment:00786"  
 164742 181183  
 /note="assembly\_fragment:03237"  
 BASE COUNT 45289 a 44875 c 44727 g 45992 t 300 others  
 ORIGIN

```
Query Match      2.7%; Score 77; DB 2; Length 181183;
Best Local Similarity 100.0%; Pred. No. 1e-28;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1979 GCAGGCACCTGTAATCCAGCTACTTGGAGGCTGAGGCAGAGAAATCGCTTGAAACCCAG 2038
      |
Db 141532 GCAGGCACCTGTAATCCAGCTACTTGGAGGCTGAGGCAGAGAAATCGCTTGAAACCCAG 141473
      |
QY 2039 GAGGTGGAGGTTGCAGT 2055
      |
Db 141472 GAGGTGGAGGTTGCAGT 141456
      |
RESULT 61
AC022281
LOCUS      186431 bp DNA linear HTG 24-JAN-2002
DEFINITION Homo sapiens chromosome 10 clone RP11-10E13, WORKING DRAFT
SEQUENCE, 9 unordered pieces.
ACCESSION AC022281
VERSION   2 (bases 1 to 186431)
KEYWORDS  HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN;
           HTGS_CANCELLED.
SOURCE    Homo sapiens
ORGANISM  Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE    1 (bases 1 to 186431)
JOURNAL  Genome Therapeutics Corporation Sequencing Center: Human Genome
AUTHORS  Smith,D.R.
REFERENCE Unpublished
AUTHORS  2 (bases 1 to 186431)
TITLE    Direct Submission
JOURNAL  Submitted (28-JAN-2000) Genome Therapeutics Corporation, 100 Beaver
          Street, Waltham, MA 02453, USA
COMMENT  On Mar 21, 2001 this sequence version replaced gi:9929643.
          -----
          Center: Genome Therapeutics Corporation
          Center code: GTC
          Web site: http://www.genomecorp.com/
          Contact: gtc-seqcenter@genomecorp.com
          -----
          Project Information
          Center project name: hgl143
          -----
          Summary Statistics
          Sequencing vector: N/A
          Chemistry: Dye-terminator Big Dye; 100% of reads
          Assembly program: Phrap; version 990315
          Consensus quality: 395389 bases at least Q40
          Consensus quality: 405067 bases at least Q30
          Consensus quality: 410525 bases at least Q20
          Insert size: 185631; sum-of-contigs
          Quality coverage: 6.4x in Q20 bases; sum-of-contigs
          -----
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 9 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *
          * 124725: contig of 124725 bp in length
          * 124726 124825: gap of unknown length
          * 124826 125924: contig of 1099 bp in length
          * 125925 126024: gap of unknown length
          * 126025 128493: contig of 2469 bp in length
          * 128494 128593: gap of unknown length
          * 128594 156515: contig of 27922 bp in length
          * 156516 156616: gap of unknown length
          * 156616 170011: contig of 13396 bp in length
          * 170011 170111: gap of unknown length
```

```
* 170112 176430: contig of 6319 bp in length
* 176431 176530: gap of unknown length
* 176531 179019: contig of 2489 bp in length
* 179020 179119: gap of unknown length
* 179120 181965: contig of 2846 bp in length
* 181966 182065: gap of unknown length
* 182066 186431: contig of 4366 bp in length.
FEATURES
      Location/Qualifiers
      source
      1..186431
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="10"
      /clone="RP11-10E13"
      /clone_lib="RFCI-11"
      1..124725
      /note="assembly_name:Contig243
      clone_end:SP6"
      misc_feature
      124826..125924
      /note="assembly_name:Contig204"
      misc_feature
      126025..128493
      /note="assembly_name:Contig233"
      misc_feature
      128594..156515
      /note="assembly_name:Contig240"
      misc_feature
      156616..170011
      /note="assembly_name:Contig239"
      misc_feature
      170112..176430
      /note="assembly_name:Contig235"
      misc_feature
      176531..179019
      /note="assembly_name:Contig231"
      misc_feature
      179120..181965
      /note="assembly_name:Contig232"
      misc_feature
      182066..186431
      /note="assembly_name:Contig236
      clone_end:T7"
BASE COUNT 48827 a 41798 c 41896 g 53062 t 848 others
ORIGIN

Query Match      2.7%; Score 77; DB 2; Length 186431;
Best Local Similarity 100.0%; Pred. No. 1e-28;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1979 GCAGGCACCTGTAATCCAGCTACTTGGAGGCTGAGGCAGAGAAATCGCTTGAAACCCAG 2038
      |
Db 132433 GCAGGCACCTGTAATCCAGCTACTTGGAGGCTGAGGCAGAGAAATCGCTTGAAACCCAG 132492
      |
QY 2039 GAGGTGGAGGTTGCAGT 2055
      |
Db 132493 GAGGTGGAGGTTGCAGT 132509.
      |
RESULT 62
AC007345
DEFINITION Homo sapiens chromosome 16 clone RP11-466N18, complete sequence.
ACCESSION AC007345
VERSION   AC007345.5 GI:11094677
KEYWORDS  HTG.
SOURCE    Homo sapiens.
ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 188863)
AUTHORS  DOE Joint Genome Institute.
TITLE    Sequencing of Human Chromosome 16
JOURNAL  Unpublished
REFERENCE 2 (bases 1 to 188863)
AUTHORS  Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E., Robinson,D.,
          Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L.,
          Bryant,J., Tesmer,J., Meincke,L., Longmire,J., White,S., Tatum,O.,
          Campbell,C., Fawcett,J., Maltbie,M., Bussod,M., Sutherland,R.,
          McMurry,K., Han,C. and Deaven,L.
TITLE    Direct Submission
JOURNAL  Submitted (21-APR-1999) Center for Human Genome Studies, DOE Joint
```

Genome Institute, Los Alamos National Laboratory, MS M888, Los Alamos, NM 87545, USA  
 3 (bases 1 to 188863)  
 DOE Joint Genome Institute.  
 Direct Submission  
 Submitted (04-NOV-2000) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
 On Nov 4, 2000 this sequence version replaced gi:9795553.  
 -----Genome Center  
 Center: Joint Genome Institute  
 Center Code: JGI  
 web site: http://www.jgi.doe.gov  
 -----

# FEATURES

source

1..188863  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="16"  
 /clone="RP11-466N18"  
 BASE COUNT 57057 a 36866 c 36556 g 58384 t  
 ORIGIN

Query Match 2.7% Score 77; DB 9; Length 188863;  
 Best Local Similarity 100.0%; Pred. No. le-28;  
 Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1979 GCAGGCACCTGTAAATCCAGCTACTTGGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAG 2038  
 Db 185044 GCAGGCACCTGTAAATCCAGCTACTTGGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAG 185103  
 QY 2039 GAGGTGGAGGTTGCAGT 2055  
 Db 185104 GAGGTGGAGGTTGCAGT 185120

# RESULT 63

AP005264/c  
 LOCUS AP005264.3  
 DEFINITION Homo sapiens genomic DNA, chromosome 18 clone:RP11-64C12, complete sequence.  
 ACCESSION AP005264  
 VERSION AP005264.3 GI:22202843  
 KEYWORDS HTG.  
 SOURCE Homo sapiens DNA, clone:RP11-64C12.  
 ORGANISM Homo sapiens

REFERENCE 1  
 AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.  
 TITLE Homo sapiens genomic DNA  
 JOURNAL Published Only in Database (2002)  
 REFERENCE 2 (bases 1 to 192539)  
 AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.  
 TITLE Direct Submission  
 JOURNAL Submitted (24-MAY-2002) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-cho,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:hattori@psc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)

On Aug 9, 2002 this sequence version replaced gi:21624263.  
 COMMENT On Aug 9, 2002 this sequence version replaced gi:21624263.  
 FEATURES  
 Location/Qualifiers  
 1..192539  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /map="18p"  
 /chromosome="18"  
 /clone="RP11-64C12"  
 BASE COUNT 51833 a 45020 c 43248 g 52438 t  
 ORIGIN

Query Match

2.7% Score 77; DB 9; Length 192539;

Best Local Similarity 100.0%; Pred. No. le-28;  
 Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1979 GCAGGCACCTGTAAATCCAGCTACTTGGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAG 2038  
 Db 95946 GCAGGCACCTGTAAATCCAGCTACTTGGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAG 95887  
 QY 2039 GAGGTGGAGGTTGCAGT 2055  
 Db 95886 GAGGTGGAGGTTGCAGT 95870

# RESULT 64

CNS01DRV

LOCUS

DEFINITION

Human chromosome 14 DNA sequence BAC R-1017G21 of library RPCI-11

from chromosome 14 of Homo sapiens (human), complete sequence.

ACCESSION AL118558

VERSION AL118558.6 GI:15485141

KEYWORDS HTG; HTGS\_ACTIVEFIN.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 194871)

AUTHORS Heilig,R., Petit,J.L., Vico,V., Dasilva,C., Robert,C., Wincker,P.,

Brottier,P., Cattolico,L., Barbe,V., Pelletier,E., Arliguene,F.,

Levy,M., Eckenberg,R., Bruls,T., Deberardinis,V., Croud,C.,

Gyapay,G., Saurin,W. and Weissbach,J.

Sequencing of the human chromosome 14

Unpublished

JOURNAL

REFERENCE 2 (bases 1 to 194871)

AUTHORS Genoscope.

TITLE Direct Submission

JOURNAL Submitted (04-SEP-2001) Genoscope - Centre National de Sequencae ;

BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr

- Web : www.genoscope.cns.fr)

On Sep 6, 2001 this sequence version replaced gi:14571656.

COMMENT ----- Genome Center

Center: Genoscope / Centre National de Sequencae

Center code: GS

Web site: http://www.genoscope.cns.fr/

Contact: Seqref@genoscope.cns.fr

-----

The following BAC sequence is oriented from the T7 to the SP6 end.

Upstream BAC (overlapping the T7 end) : C-2017C7

Downstream BAC (overlapping the SP6 end) : R-114H15 (AC AL14221)

----- Summary Statistics

Assembly program: Phrap; version 2.0

Quality coverage: 9.00x in 920 bases; sum-of-contigs

-----

Overall quality chart :

Range : bases

0 : 9

1 - 9 : 1

10 - 19 : 36

20 - 29 : 176

30 - 39 : 2561

40 - 49 : 6688

50 - 59 : 12185

60 - 69 : 30197

70 - 79 : 70225

80 - 89 : 72822

90 - 99 : 72822

Percentage of bases with a quality value >= 40 : 99 %.

Location/Qualifiers

1..194871

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="14"

/clone="R-1017G21"

FEATURES

source





misc_feature	/note="81% identity -- 277249" 21027..21097	misc_feature	complement(37454..37580) /note="84% identity" /db_xref="dbEST:T62489"
repeat_region	complement(24077..24350) /note="U47924"	repeat_region	complement(37828..38086) /note="270040"
repeat_region	/rpt_family="Alu" complement(24301..24492) /note="HSAL000410"	repeat_region	/rpt_family="MER4" complement(41427..41642) /rpt_family="MLT2B2"
repeat_region	24633..24901 /rpt_family="Alu"	repeat_region	complement(42308..42622) /rpt_family="MLT2B2"
repeat_region	/rpt_family="Alu" complement(24992..25278) /note="273359"	repeat_region	43022..43522 /rpt_family="MLT2B2"
repeat_region	/rpt_family="Alu" complement(25515..25599) /note="MALR00365"	misc_feature	43881..44000 /rpt_family="Tigger2"
repeat_region	/rpt_family="THE1" complement(25777..25901) /note="MALR00322"	misc_feature	44984..45116 /note="GRAIL 2 good exon ; ORF; frame 2"
repeat_region	/rpt_family="MSTC" complement(26808..27050) /note="X66298"	misc_feature	complement(46422..46478) /note="GRAIL 2 good exon ; ORF; frame 0"
repeat_region	/rpt_family="Alu" complement(27009..27125) /note="HSAL004729"	repeat_region	complement(46801..47121) /note="HSAL005507"
misc_feature	complement(27361..27462) /note="GRAIL 2 excellent exon ; ORF; frame 2"	misc_feature	/rpt_family="Alu" complement(47511..47779) /note="GRAIL 2 good exon ; ORF; frame 1"
misc_feature	complement(27482..27721) /note="GRAIL 2 good exon ; ORF; frame 2"	repeat_region	complement(48123..48196) /note="MALR00186"
misc_feature	30322..31155 /rpt_family="MER33"	misc_feature	/rpt_family="MLTc" complement(49320..49497) /note="GRAIL 2 excellent exon ; ORF; frame 2"
gene	31815..31991 /note="GRAIL 2 good exon ; ORF; frame 1"	misc_feature	complement(49334..50215) /note="GRAIL 2 good exon ; ORF; frame 2"
mRNA	/gene="VIP2R" complement(<31907..34394) /note="VIP2R"	repeat_region	complement(51833..51901) /note="GRAIL 2 good exon ; ORF; frame 2"
misc_feature	/product="vasoactive intestinal polypeptide receptor 2 precursor" complement(31907..32006) /gene="VIP2R"	repeat_region	complement(52214..52500) /note="268162"
CDS	/note="99% identity -- L40764" complement(join(<31908..32006,34182..34234)) /gene="VIP2R"	repeat_region	/rpt_family="Alu" 54960..55032 /rpt_family="L1MC2"
	/codon_start=3 /product="vasoactive intestinal polypeptide receptor 2 precursor" /db_xref="GI:3004859"	repeat_region	55462..55537 /rpt_family="L1ME2"
misc_feature	/translation="MRTLLPALLTCWLLAPVNSIHPECRFHLEIOEEETKCAELLRS QTEKKH" complement(33411..33597) /gene="VIP2R"	repeat_region	57335..58234 /rpt_family="L1"
misc_feature	/note="100% identity -- 259087" complement(34182..34232) /gene="VIP2R"	repeat_region	complement(58350..58466) /rpt_family="MER44"
misc_feature	/note="100% identity -- L40764" complement(34200..34394) /gene="VIP2R"	misc_feature	/rpt_family="L1MA9" complement(58757..59020) /note="GRAIL 2 good exon ; ORF; frame 0"
repeat_region	34432..34694 /rpt_family="TAR1"	repeat_region	62145..62345 /note="GRAIL 2 good exon ; ORF; frame 2"
misc_feature	35000..35058 /note="GRAIL 2 excellent exon ; ORF; frame 2"	repeat_region	complement(62518..64744) /rpt_family="Tigger1"
repeat_region	complement(35325..35445) /rpt_family="MER42"	misc_feature	complement(66883..67070) /note="GRAIL 2 excellent exon ; ORF; frame 0"
repeat_region	complement(36238..36296) /note="HSAL004342"	repeat_region	complement(67293..67336) /note="GRAIL 2 good exon ; ORF; frame 1"
repeat_region	/rpt_family="Alu" complement(37248..37455) /note="MER4_00001"	misc_feature	67421..67746 /rpt_family="Alu"
	/rpt_family="MER4"	misc_feature	complement(68255..68320)

Query Match 2.7% Score 77; DB 9: Length 225412;  
Best Local Similarity 100.0%; Pred. No. le-28;  
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1979	GCAGGCACCTGTATCCAGCTACTTGGGAGGCTGAGGCAGGAGATCGTTGAATCCAG	2038
Db	124031	GCAGGCACCTGTATCCAGCTACTTGGGAGGCTGAGGCAGGAGATCGTTGAATCCAG	123972
Qy	2039	GAGGTGAGGTTGCAGT	2055
Db	123971	GAGGTGAGGTTGCAGT	123955
RESULT 66			





\* 14895 14994: gap of 100 bp  
\* 14995 15703: contig of 709 bp in length  
\* 15704 15803: gap of 100 bp  
\* 15804 16511: contig of 708 bp in length  
\* 16512 16611: gap of 100 bp  
\* 16612 17284: contig of 673 bp in length  
\* 17285 17384: gap of 100 bp  
\* 17385 18098: contig of 714 bp in length  
\* 18099 18198: gap of 100 bp  
\* 18199 18905: contig of 707 bp in length  
\* 18906 19005: gap of 100 bp  
\* 19006 19710: contig of 705 bp in length  
\* 19711 19810: gap of 100 bp  
\* 19811 20486: contig of 676 bp in length  
\* 20487 20586: gap of 100 bp  
\* 20587 21266: contig of 680 bp in length  
\* 21267 21366: gap of 100 bp  
\* 21367 22034: contig of 668 bp in length  
\* 22035 22134: gap of 100 bp  
\* 22135 22814: contig of 680 bp in length  
\* 22815 22914: gap of 100 bp  
\* 22915 23594: contig of 680 bp in length  
\* 23595 23694: gap of 100 bp  
\* 23695 24383: contig of 689 bp in length  
\* 24384 24483: gap of 100 bp  
\* 24484 25187: contig of 704 bp in length  
\* 25188 25287: gap of 100 bp  
\* 25288 25989: contig of 702 bp in length  
\* 25990 26089: gap of 100 bp  
\* 26090 26798: contig of 709 bp in length  
\* 26799 26998: gap of 100 bp  
\* 26999 27607: contig of 709 bp in length  
\* 27608 27707: gap of 100 bp  
\* 27708 28387: contig of 680 bp in length  
\* 28388 28487: gap of 100 bp  
\* 28488 29152: contig of 665 bp in length  
\* 29153 29252: gap of 100 bp  
\* 29253 29944: contig of 692 bp in length  
\* 29945 30044: gap of 100 bp  
\* 30045 30733: contig of 689 bp in length  
\* 30734 30833: gap of 100 bp  
\* 30834 31548: contig of 715 bp in length  
\* 31549 31648: gap of 100 bp  
\* 31649 32356: contig of 708 bp in length  
\* 32357 32456: gap of 100 bp  
\* 32457 33164: contig of 708 bp in length  
\* 33165 33264: gap of 100 bp  
\* 33265 33938: contig of 674 bp in length  
\* 33939 34038: gap of 100 bp  
\* 34039 34751: contig of 713 bp in length  
\* 34752 34851: gap of 100 bp  
\* 34852 35561: contig of 710 bp in length  
\* 35562 35661: gap of 100 bp  
\* 35662 36313: contig of 652 bp in length  
\* 36314 36413: gap of 100 bp  
\* 36414 37093: contig of 680 bp in length  
\* 37094 37193: gap of 100 bp  
\* 37194 37868: contig of 675 bp in length  
\* 37869 37968: gap of 100 bp  
\* 37969 38647: contig of 679 bp in length  
\* 38648 38747: gap of 100 bp  
\* 38748 39435: contig of 688 bp in length  
\* 39436 39535: gap of 100 bp  
\* 39536 40233: contig of 698 bp in length  
\* 40234 40333: gap of 100 bp  
\* 40334 41023: contig of 690 bp in length  
\* 41024 41123: gap of 100 bp  
\* 41124 41824: contig of 701 bp in length  
\* 41825 41924: gap of 100 bp  
\* 41925 42607: contig of 683 bp in length  
\* 42608 42707: gap of 100 bp  
\* 42708 43411: contig of 704 bp in length  
\* 43412 43511: gap of 100 bp

\* 43512 44217: contig of 706 bp in length  
\* 44218 44317: gap of 100 bp  
\* 44318 45001: contig of 684 bp in length  
\* 45002 45101: gap of 100 bp  
\* 45102 45783: contig of 682 bp in length  
\* 45784 45883: gap of 100 bp  
\* 45884 46575: contig of 692 bp in length  
\* 46576 46675: gap of 100 bp  
\* 46676 47368: contig of 693 bp in length  
\* 47369 47468: gap of 100 bp  
\* 47469 48149: contig of 681 bp in length  
\* 48150 48249: gap of 100 bp  
\* 48250 48949: contig of 700 bp in length  
\* 48950 49049: gap of 100 bp  
\* 49050 49756: contig of 707 bp in length  
\* 49757 49856: gap of 100 bp  
\* 49857 50557: contig of 701 bp in length  
\* 50558 50657: gap of 100 bp  
\* 50658 51340: contig of 683 bp in length  
\* 51341 51440: gap of 100 bp  
\* 51441 52140: contig of 700 bp in length  
\* 52141 52240: gap of 100 bp  
\* 52241 52932: contig of 692 bp in length  
\* 52933 53032: gap of 100 bp  
\* 53033 53719: contig of 687 bp in length  
\* 53720 53819: gap of 100 bp  
\* 53820 54503: contig of 684 bp in length  
\* 54504 54603: gap of 100 bp  
\* 54604 55294: contig of 691 bp in length

Query Match 2.6%; Score 74; DB 2: Length 66452;  
Best Local Similarity 100.0%; Pred. No. 3.8e-27;  
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1982 GGCACCTGTATCCAGCTACTTGGAGGCTGAGCAGAGAAATCGTTGAACCCAGGAG 2041  
|||||  
Db 9693 GGCACCTGTATCCAGCTACTTGGAGGCTGAGCAGAGAAATCGTTGAACCCAGGAG 9752  
|||||

QY 2042 GTGGAGTTGCAGT 2055  
|||||

Db 9753 GTGGAGTTGCAGT 9766  
|||||

RESULT 69  
AC008754/c  
LOCUS AC008754 66792 bp DNA linear PRI 28-FEB-2001  
DEFINITION Homo sapiens chromosome 19 clone CTD-3023J11, complete sequence.  
ACCESSION AC008754  
VERSION AC008754.8 GI:13162498  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 66792)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 66792)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
REFERENCE 3 (bases 1 to 66792)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Submitted (28-FEB-2001) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA  
COMMENT On Feb 28, 2001 this sequence version replaced gi:9256040.  
Draft Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov  
Finishing Completed at Stanford Human Genome Center  
www-shgc.stanford.edu

Quality: Phrap Quality >=40 99.6% of Sequence;  
Estimated Total Number of Errors is 0.2.

# FEATURES

```
Source
1..66792
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/clone="CTD-3023J11"
BASE COUNT 16364 a 17164 c 17596 g 15668 t
ORIGIN
Query Watch 2.6%; Score 74; DB 9; Length 66792;
Best Local Similarity 100.0%; Pred. No. 3.8e-27;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1982 GGCACCTGTAATCCAGCTACTTGGAGGCTGAGCAGGAGTAATCGCTTGAACCCAGGAG 2041
|||||
DB 15755 GGCACCTGTAATCCAGCTACTTGGAGGCTGAGCAGGAGTAATCGCTTGAACCCAGGAG 15696
|||||
QY 2042 GTGGAGGTTGCAGT 2055
|||||
DB 15695 GTGGAGGTTGCAGT 15682
|||||
```

# RESULT 70

```
AC092579
LOCUS Homo sapiens BAC clone RP11-15K19 from 7, complete sequence.
DEFINITION AC092579 AC011922
AC092579.3 GI:18653765
VERSION
KEYWORDS
SOURCE
Homo sapiens.
```

# ORGANISM

```
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 118958)
Sulston, J.E. and Waterston, R.H.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
99063792
9847074
```

# REFERENCE

```
2 (bases 1 to 118958)
Cordum, H., Elliott, G., Spalding, L. and Phillips, A.
The sequence of Homo sapiens BAC clone RP11-15K19
Unpublished (2001)
3 (bases 1 to 118958)
Waterston, R.H.
Direct Submission
Submitted (19-JUL-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 118958)
Waterston, R.H.
Direct Submission
Submitted (13-FEB-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 118958)
Waterston, R.H.
Direct Submission
Submitted (21-FEB-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Feb 13, 2002 this sequence version replaced gi:18425307.
```

# REFERENCE

```
AUTHORS
TITLE
JOURNAL
Center: Washington University Genome Sequencing Center
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.edu
Center project name: H_NH0015K19
Drafting Center: WIBR
-----
```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred > 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

# MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

# SOURCE INFORMATION:

The RPC1-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, H., Frenken, E., Taten, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org> VECTOR: pBAC3.6

# NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-419M24, 2000 bp overlap; the clone sequenced to the right is RP11-745J15. Actual start of this clone is at base position 160861 of RP11-419M24; actual end is at base position 118958 of RP11-15K19.

Data from AC092101 was used to finish AC092579.

The sequence of AC011922 has been incorporated into AC092579.

```
Location/Qualifiers
1..118958
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7"
/clone="RP11-15K19"
/clone_lib="RPC1-11"
1..139
/rpt_family "Alu"
193..497
/rpt_family "Alu"
228..245
/misc_feature
553..864
/rpt_family "Alu"
821..1090
/misc_feature
918..993
/rpt_family "MER1_type?"
1053..1499
/misc_feature
1646..1957
/rpt_family "Alu"
2002..2270
/rpt_family "MaLR"
2271..3282
/rpt_family "ERVK"
2920..2928
/misc_feature
3283..3428
/rpt_family "Alu"
```

```
/rpt_family="MaLR"
3503..3812
/rpt_family="Alu"
4056
/Note="match to EST AW903301 (NID:g8067506)"
4086..4447
/Note="match to EST AW903301 (NID:g8067506)"
4417..4720
/rpt_family="Alu"
4574..4575
/Note="match to EST AW903301 (NID:g8067506)"
4790..5093
/rpt_family="Alu"
5248..5250
/Note="match to EST N68717 (NID:g1224878) za20a06.sl"
5376..5830
/Note="match to EST AI492948 (NID:g4393951) qz47a04.xl"
5933..6286
/rpt_family="MaLR"
6328..6602
/rpt_family="Alu"
6340..6369
/Note="match to EST N68717 (NID:g1224878) za20a06.sl"
6988..7281
/rpt_family="Alu"
7282..7543
/rpt_family="Alu"
7443..7469
/Note="match to EST BF668855 (NID:g11942750)"
8243..8540
/rpt_family="Alu"
8541..8900
/rpt_family="L1"
8898..8915
/Note="similar to Homo sapiens EST B1856928
(NID:g19997675)"
8901..9212
/rpt_family="Alu"
9213..9329
/rpt_family="L1"
9330..9601
/rpt_family="Alu"
9602..9904
/rpt_family="L1"
9905..10191
/rpt_family="Alu"
10192..10412
/rpt_family="L1"
10413..10718
/rpt_family="Alu"
10727..11036
/rpt_family="Alu"
11058..12390
/Note="CpG island (GC=60.2, o/e=0.80, #CpGs=97)"
11260..11309
/Note="match to EST B1458323 (NID:g15248979)"
11280..11309
/Note="match to EST BE827799 (NID:g10260177)"
11281..11309
/Note="match to EST BE439838 (NID:g9439449)"
11297..11309
/Note="match to EST AA358556 (NID:g2010873)"
11341..11655
/Note="match to EST BF237438 (NID:g1151356)"
11678..11750
/rpt_family="MIR"
11830..12121
/rpt_family="Alu"
12044..12049
/Note="match to EST N55773 (NID:g1198621)"
12129..12428
/rpt_family="Alu"
12550..12697

/rpt_family="MaLR"
12558..12697
/Note="match to EST AL539603 (NID:g12868971)"
12558..12697
/Note="match to EST AA358556 (NID:g2010873)"
12558..12697
/Note="match to EST BE439838 (NID:g9439449)"
12558..12697
/Note="match to EST BE827799 (NID:g10260177)"

Query Match 2.6% Score 74: DB 9: Length 118958;
Best Local Similarity 100.0%; Pred.No. 3.7e-27;
Matches 74: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1982 GGCACCTGTAATCCAGCTACTTGGGAGGCTGAGCAGGAGAGTAATCGTTGAACCCAGGAG 2041
Db 60317 GGCACCTGTAATCCAGCTACTTGGGAGGCTGAGCAGGAGAGTAATCGTTGAACCCAGGAG 60376
QY 2042 GTGGAGGTTGCAGT 2055
Db 60377 GTGGAGGTTGCAGT 60390

RESULT 71
AF214635 158213 bp DNA linear HTG 26-JUL-2002
LOCUS Homo sapiens chromosome 8 clone XX-CTB788C1, WORKING DRAFT
DEFINITION SEQUENCE, 9 unordered pieces.
ACCESSION AF214635
VERSION AF214635.3 GI:14280180
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_CANCELLED.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 158213)
Reichwald,K., Baumgart,C., Blechschmidt,K., Dette,M., Jahn,N.,
Lehmann,R., Menzel,U., Polley,A., Schilhabel,M.B., Schudy,A.,
Siddiqui,R., Taudien,S., Wen,G., Korenberg,J.R., Rosenthal,A. and
Platzter,M.
Chromosome 8 genomic sequence
Unpublished
2 (bases 1 to 158213)
Reichwald,K., Blechschmidt,K., Menzel,U., Baumgart,C., Dette,M.,
Jahn,N., Schilhabel,M., Korenberg,J.R. and Rosenthal,A.
Direct Submission
TITLE
JOURNAL
COMMENT
On Jun 2, 2001 this sequence version replaced gi:8152159.
----- Genome Center
Center: Institute of Molecular Biotechnology
Center code: IMB
Web site: http://genome.imb-jena.de/
Contact: gscj-submit@genome.imb-jena.de
----- Project Information
Center project name: H195
Center clone name: XX-CTB788C1
----- Summary Statistics
Sequencing vector: M13; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 146316 bases at least Q40
Consensus quality: 151479 bases at least Q30
Consensus quality: 154827 bases at least Q20
Quality coverage: 5.51 x in Q20 bases; sum-of-contigs
-----
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.
```

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 9 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\*  
\* 3530: contig of 3530 bp in length  
\* 3531 3530: gap of unknown length  
\* 3631 5131: contig of 1501 bp in length  
\* 5132 5231: gap of unknown length  
\* 5232 13445: contig of 8214 bp in length  
\* 13446 13545: gap of unknown length  
\* 13546 31026: contig of 17481 bp in length  
\* 31027 31126: gap of unknown length  
\* 31127 51905: contig of 20779 bp in length  
\* 51906 52005: gap of unknown length  
\* 52006 81566: contig of 29561 bp in length  
\* 81567 81666: gap of unknown length  
\* 81667 111349: contig of 29683 bp in length  
\* 111350 111449: gap of unknown length  
\* 111450 154738: contig of 43289 bp in length  
\* 154739 154838: gap of unknown length  
\* 154839 158213: contig of 3375 bp in length.  
FEATURES  
    Location/Qualifiers  
        1..158213  
            /organism="Homo sapiens"  
            /db\_xref="taxon:9606"  
            /chromosome="8"  
            /clone="XX-CTB788C1"  
        1..4  
            /Note="assembly\_fragment"  
            clone\_end:SP6  
            vector\_side:left  
            158210..158213  
            /Note="assembly\_fragment"  
            clone\_end:T7  
            vector\_side:right  
BASE COUNT 42142 a 33754 c 33773 g 47744 t 800 others  
ORIGIN  
  
Query Match 2.6%; Score 74; DB 2; Length 158213;  
Best Local Similarity 100.0%; Pred. No. 3.7e-27;  
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1986 CCGTGAATCCCGACTGTGGAGGCTGAGGAGGAGGAGTGCCTTGAACCCAGGAGGTGG 2045  
          |||||  
Db 137693 CCGTGAATCCCGACTGTGGAGGCTGAGGAGGAGGAGTGCCTTGAACCCAGGAGGTGG 137752  
          |||||  
QY 2046 AGGTTGCAGTAAGC 2059  
          |||||  
Db 137753 AGGTTGCAGTAAGC 137766  
          |||||  
  
RESULT 72  
AC105081/c 158605 bp DNA linear PRI 27-APR-2002  
LOCUS Homo sapiens chromosome 8, clone CTC-788C1, complete sequence.  
DEFINITION AC105081  
ACCESSION AC105081  
VERSION AC105081.4 GI:20258535  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 158605)  
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
TITLE Homo sapiens chromosome 8, clone CTC-788C1  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 158605)  
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,

Anderson,S., Barna,N., Bastien,V., Boquslavkiy,L., Boukhgalter,B.,  
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,  
Choepe,Y., Colangelo,M., Collins,S., Collamore,A., Cook,A.,  
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,  
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,  
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,  
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,  
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,  
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,  
McCarthy,M., McEwan,P., McKernan,K., Meidrim,J., Menus,L.,  
Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,  
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,  
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,  
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,  
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,  
Seaman,S., Severi,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
Submitted (23-DEC-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 158605)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boquslavkiy,L.,  
Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,  
Chazaro,B., Choepe,Y., Colangelo,M., Collins,S., Collamore,A.,  
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,  
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,  
Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,  
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,  
McCarthy,M., McEwan,P., McKernan,K., Meidrim,J., Menus,L.,  
Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,  
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,  
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,  
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,  
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,  
Seaman,S., Severi,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (21-APR-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
4 (bases 1 to 158605)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boquslavkiy,L.,  
Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,  
Chazaro,B., Choepe,Y., Colangelo,M., Collins,S., Collamore,A.,  
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,  
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,  
Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,  
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,  
McCarthy,M., McEwan,P., McKernan,K., Meidrim,J., Menus,L.,  
Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,  
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,  
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,  
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,  
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,  
Seaman,S., Severi,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (21-APR-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
4 (bases 1 to 158605)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boquslavkiy,L.,  
Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,  
Chazaro,B., Choepe,Y., Colangelo,M., Collins,S., Collamore,A.,  
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,  
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,  
Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,  
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,  
McCarthy,M., McEwan,P., McKernan,K., Meidrim,J., Menus,L.,  
Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,  
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,  
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,  
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,  
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,  
Seaman,S., Severi,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.



TITLE Direct Submission  
 JOURNAL Submitted (27-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 COMMENT On Apr 21, 2002 this sequence version replaced gi:20128267.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L21405  
 Center clone name: 788\_C\_1  
 -----

## FEATURES

Source

Location/Qualifiers

1..158605  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="8"  
 /map="8"  
 /clone="CTC-788C1"  
 /clone\_lib="CIRC Human BAC"  
 complement(339..543)  
 /rpt\_family="MLT1k"  
 545..709  
 /rpt\_family="GA-rich"  
 complement(1479..1582)  
 /rpt\_family="MIR"  
 1637..1664  
 /rpt\_family="AT-rich"  
 complement(4850..5036)  
 /rpt\_family="MER5A"  
 5091..5119  
 /rpt\_family="(A)n"  
 5290..5447  
 /rpt\_family="LIMA3"  
 complement(5645..5710)  
 /rpt\_family="MER91A"  
 complement(5740..5803)  
 /rpt\_family="MER91A"  
 6092..6264  
 /rpt\_family="MLT1h"  
 6360..7020  
 /rpt\_family="MER67B"  
 7101..7221  
 /rpt\_family="MLT1h1"  
 complement(7495..7856)  
 /rpt\_family="MLT1A1"  
 complement(8994..9054)  
 /rpt\_family="L2"  
 complement(9121..9220)  
 /rpt\_family="MIR"  
 complement(9353..9430)  
 /rpt\_family="L2"  
 complement(9454..9527)  
 /rpt\_family="MIR"  
 10413..10430  
 /rpt\_family="MIR"  
 complement(10431..10651)  
 /rpt\_family="MER58A"  
 10652..10879  
 /rpt\_family="MIR"  
 12388..12447  
 /rpt\_family="L2"  
 12605..12827  
 /rpt\_family="MIR"  
 13849..14151  
 /rpt\_family="MSTD"  
 14152..14451  
 /rpt\_family="AluSx"  
 14452..14520

repeat\_region /rpt\_family="MSTD"  
 14546..14601  
 /rpt\_family="HAL1"  
 repeat\_region 14603..15150  
 /rpt\_family="HAL1"  
 repeat\_region 15148..15313  
 /rpt\_family="LIMB2"  
 repeat\_region 15314..15616  
 /rpt\_family="AluSx"  
 repeat\_region 15617..15661  
 /rpt\_family="LIMB2"  
 repeat\_region 15662..15693  
 /rpt\_family="(TG)n"  
 repeat\_region 15694..15729  
 /rpt\_family="LIMB2"  
 repeat\_region 15730..15759  
 /rpt\_family="(GTGTG)n"  
 repeat\_region 15760..15849  
 /rpt\_family="LIMB2"  
 repeat\_region 15850..15974  
 /rpt\_family="AluY"  
 repeat\_region 15975..16001  
 /rpt\_family="(CAAAA)n"  
 repeat\_region 16002..16179  
 /rpt\_family="AluY"  
 repeat\_region 16180..16345  
 /rpt\_family="LIMB2"  
 repeat\_region 16346..16750  
 /rpt\_family="LIPA2"  
 repeat\_region 16819..16940  
 /rpt\_family="LIMB2"  
 repeat\_region 17420..17513  
 /rpt\_family="LIMB2"  
 repeat\_region complement(17575..17683)  
 /rpt\_family="LIMC4"  
 repeat\_region 17812..17896  
 /rpt\_family="LIMB2"  
 repeat\_region 18004..18034  
 /rpt\_family="AT-rich"

Query Match 2.6%; Score 74; DB 9; Length 158605;

Best Local Similarity 100.0%; Pred.No. 3.7e-27;

Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1986 CCTGTAATCCCGAGCTACTGGGAGGCTGAGCAGGAGAGTAATCGTTGAACCCAGGAGGTGG 2045

Db 129206 CCTGTAATCCCGAGCTACTGGGAGGCTGAGCAGGAGAGTAATCGTTGAACCCAGGAGGTGG 129147

QY 2046 AGGTTGCAGTAAGC 2059

Db 129146 AGGTTGCAGTAAGC 129133

## RESULT 73

AC068793/c

LOCUS

DEFINITION Homo sapiens chromosome 12 clone RP11-414A12, WORKING DRAFT

SEQUENCE, 8 unordered pieces.

ACCESSION AC068793

VERSION AC068793.19 GI:21431064

KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 160371)

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,

Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,

Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,

Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,

Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,

Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,

Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,



Weinstock, I.R., Williamson, A., Worley, K., Wren, J., Wrensford, G.,  
Xiang, A.M., Yang, R., Yu, W., Zhou, X., Kucherlapati, R., Nelson, D. and  
Gibbs, R.

# TITLE

## REFERENCE

2. (bases 1 to 162200)

# AUTHORS

## JOURNAL

Submitted (21-OCT-1999) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

3. (bases 1 to 162200)

# AUTHORS

## JOURNAL

Submitted (31-MAR-2000) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

4. (bases 1 to 162200)

# AUTHORS

## JOURNAL

Submitted (12-AUG-2000) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

On Mar 31, 2000 this sequence version replaced gi:7211838.

INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email

gc-help@bcm.tmc.edu

## COMMENT

CLONE LENGTH: This sequence does not necessarily represent the  
entire insert of this clone. Overlapping regions of clones are only  
sequenced and submitted once, so the sequence for the remainder of  
the insert may be found in the record for the adjacent clones.  
Overlapping clones are noted at the beginning and end of the  
Features listing.

## ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches  
of a local database that includes entries from dbSTS, GDB, and  
local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green,  
unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST  
(Nuc. Acids Res 25:3389-3402) similarity (expect < 1e-34) to the  
EST and cDNA sequences. Genes demonstrate at least 2 exons  
flanked by consensus splice sites that maintained sequence  
continuity across the splice junctions. Sequences that are not  
identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum  
standard of double strand coverage with a minimum of 2 clones and 2  
reads with no ambiguities or 2 chemistries with a minimum of 2  
clones and 3 reads with no ambiguities. If the sequence quality for  
a region does not meet this standard, it will be indicated in the  
annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality  
standards - estimated error rate less than 1 per 10,000 bases.  
Reports of lowest quality individual bases and measures of base  
quality are listed below. Description of the metrics can be found  
at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

## QUALSTAT-REPORT-----

----- Summary Statistics -----  
Contig length: 162200  
Phrap values in estimate: 156878  
Average error rate (BCM-Phrap estimate): 0.000360119  
Fraction of Phrap values less than 40 : 0.0298257  
Number of consensus changing edits: 53  
Number of N's in consensus : 0

Position	Consensus changing edits	Edited:Context
27	Original+Context	atagtactg(t)ttgtgtgca
18532		tattctctatg(t)ttttttttt
18533		attctctatg(t)ttttttttt
18534		ttctctatg(t)ttttttttt
18561		ttttttttt(g)ganacnagt
18564		tttttttga(g)acagatctt
18568		tttggacac(g)agctctgtc
18714		agctgattt(c)ttttgtatt
18730		gtatttttag(t)ggagacggg
18734		tttttaggga(n)acgctgttc
18754		cacctgttta(n)ccagatgtt
18781		ctctgacct(n)gngatccgc
18783		ctctgacct(n)gatccgcga
18795		atccgacct(n)nggcccacc
18797		ccgacctat(n)ggcccccaa
22643		tacctctttc(n)acacagcat
22705		ctttctatt(n)tacaanccg
22712		attttacaaa(n)ccgcatgtt
22760		tggttacacc(n)ncagacggg
2761		gggtacacn(n)ccagacggg
3498		aactctggag(n)tnaagcaca
3500		ctctggagnt(n)aaagcaca
35410		gcccctaagt(n)ctgtctgcc
55354		gtttattga(n)aaagcagg
55718		cttttagatg(g)attttctct
55775		gagaattttc(a)accataaac
55836		tagctgacct(n)attatgctt
55992		cacaagatc(n)ctnncttcat
55995		aaagtacct(n)ntctcatitt
55996		aagtacnctn(n)cttcatttt
56268		tggtgaacac(n)ngtagtccca
56269		ggtgaacacn(n)tagtccca
67116		gcccagcct(n)taagtaacc
102871		agagatttt(n)ttccacttca
102967		tttttttaa(a)aaacagatc
119358		ccaagtatac(a)cttacttag
131009		tgaaaaaat(c)caataaaga
133724		tgccactgca(n)tgactccag
138630		aggagctga(n)gcagcgnnn
138635		gctgagcag(n)cggnnnctn
138638		gagcagcag(n)nnntcnaga
138639		angcagcag(n)nnntcnaga
138640		ngcagcag(n)nttcnagacc
138641		gcagcagcag(n)nttcnagacc
138642		cagcagcag(n)tcnagaccag
138645		ncggnnnctc(n)agaccagcct
139044		gtctctttaa(n)aaagaaaga
141300		taaaagcaaca(n)taaaagaaa
151715		ctggagtga(n)tggtngatc
151721		ctcgattgtg(n)gatctcgct
151751		ctccgctctc(n)aggttcaagc
155920		agtgctggga(n)nacagcgtg
155921		gtgctgggat(t)acagcgtga

## ----- Distribution of Quality < 40 Bases -----

# bases	5	10	15	20	25	30	35	40
1000							*	*
900							*	*
800							*	*
700							*	*
600							*	*
500							*	*
400					*		*	*
300					*		*	*
200			*	*	*		*	*
100	*	*	*	*	*	*	*	*
0	*	*	*	*	*	*	*	*

Phrap Value Range

```

-----
Version: 1.01 oxfo
Location/Qualifiers
1. .162200
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-433D24"
complement(280..585)
/rpt_family="AluS9"
complement(1341..1341)
/rpt_family="L1MA6"
1342..1641
/rpt_family="AluSx"
complement(1642..1669)
/rpt_family="L1MA6"
1670..1698
/rpt_family="(CA)n"
complement(1699..1882)
/rpt_family="L1MA6"
4674..4845
/rpt_family="MER105"
5070..5848
/rpt_family="L1PA3"

Query Match          2.6%  Score 74;  DB 9;  Length 162200;
Best Local Similarity 100.0%;  Pred. No. 3.7e-27;
Matches 74;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy  1982  GGCACCTGTATCCAGCTACTTGGGAGGCTGAGGAGGAGAAATCGCTTGAACCCAGGAG 2041
      |||||||
Db  155789 GGCACCTGTATCCAGCTACTTGGGAGGCTGAGGAGGAGAAATCGCTTGAACCCAGGAG 155730
      |||||||

Qy  2042  GTGGAGGTTGCAGT 2055
      |||||||
Db  155729 GTGGAGGTTGCAGT 155716

RESULT 75
AC073548          167722 bp  DNA  linear  PRI 31-JUL-2002
LOCUS
DEFINITION
AC073548
AC073548.5  GI:22024556
VERSION
AC073548.5  GI:22024556
KEYWORDS
HTG.
SOURCE
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
DOE Joint Genome Institute and Stanford Human Genome Center.

REFERENCE
1 (bases 1 to 167722)
AUTHORS
DOE Joint Genome Institute
TITLE
Direct Submission
JOURNAL
Unpublished
2 (bases 1 to 167722)
AUTHORS
DOE Joint Genome Institute
TITLE
Direct Submission
JOURNAL
Submitted (22-JUN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 167722)
AUTHORS
DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE
Direct Submission
JOURNAL
Submitted (31-JUL-2002) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Jul 31, 2002 this sequence version replaced gi:15022024.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.5% of Sequence;
Estimated Total Number of Errors is 0.7.
NOTE: Forced join 65004.
-----
Location/Qualifiers

```

```

source
1. .167722
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/clone="RP11-43N16"
65004
/note:"NOTE: Forced join 65004"
BASE COUNT 38329 a 46532 c 43717 g 39144 t
ORIGIN

Query Match          2.6%  Score 74;  DB 9;  Length 167722;
Best Local Similarity 100.0%;  Pred. No. 3.7e-27;
Matches 74;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy  1982  GGCACCTGTATCCAGCTACTTGGGAGGCTGAGGAGGAGAAATCGCTTGAACCCAGGAG 2041
      |||||||
Db  32512  GGCACCTGTATCCAGCTACTTGGGAGGCTGAGGAGGAGAAATCGCTTGAACCCAGGAG 32571
      |||||||

Qy  2042  GTGGAGGTTGCAGT 2055
      |||||||
Db  32572  GTGGAGGTTGCAGT 32585

RESULT 76
AC018633/c
LOCUS
DEFINITION
AC018633
AC018633.2  GI:6729063
VERSION
AC018633.2  GI:6729063
KEYWORDS
HTG.
SOURCE
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Kaul,R.K., Yu,J., Wong,G.K.-S., Magness,C.L., Green,E.D., Green,P.
and Olson,M.V.
Large-scale MCD Mapping and Sequencing of Human Chromosome 7
Unpublished
2 (bases 1 to 169234)
AUTHORS
Bubb,K.L., Desmarais,C.L., Ramsey,S.A. and Hubley,R.M.
TITLE
Direct Submission
JOURNAL
Submitted (15-DEC-1999) Human Genome Center, University of
Washington, Box 352145, Seattle, WA 98195, USA
3 (bases 1 to 169234)
AUTHORS
Kaul,R.K. and Richards,B.K.
TITLE
Direct Submission
JOURNAL
Submitted (21-JAN-2000) Human Genome Center, University of
Washington, Box 352145, Seattle, WA 98195, USA
On Jan 21, 2000 this sequence version replaced gi:6579285.

----- Genome Center:
Center: University of Washington Genome Center
Center code: UWGC
Web site: http://genome.washington.edu
Contact: uwgchelp@u.washington.edu
----- Project Information
Center project name: chr-7
Center clone name: djs1 (RP11-16G1)
----- Summary Statistics
Sequencing vector: M13; 100% of reads
Chemistry: Dye-primer Bodipy; 90% of reads
Chemistry: Dye-terminator Big Dye; 10% of reads
Assembly program: Phrap; version 0.990319
Insert size: 169 234; sum-of-contigs
Quality coverage: 11.1X in Q20 bases; sum-of-contigs
-----
Overlapping Sequences:
5' : UWGC:djs21
3' : UWGC:djs29
-----
Sequence Quality Assessment:

```

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.

Double stranded (DS) coverage: 96.4%  
DS or two chemistry coverage: 97.1%  
Single stranded regions: 10

## Sequence Validation:

This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-600 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

BglII	FP	Seq	EcoRI	FP	Seq	NsiI
-----	-----	-----	-----	-----	-----	-----
2055.00	2061.00	8954.00	8687.00	11080.00	11109.00	-----
5487.00	5270.00	7648.00	7660.00	4611.00	4562.00	-----
1913.00	1893.00	7648.00	7556.00	11080.00	11183.00	-----
1508.00	1491.00	838.00	801.00	1377.00	1369.00	-----
12901.00	13322.00	4399.00	4303.00	2193.00	2193.00	-----
1004.00	992.00	4140.00	4094.00	3735.00	3692.00	-----
529.00	503.00	660.00	643.00	614.00	599.00	-----
10358.00	10526.00	3053.00	3100.00	3918.00	3851.00	-----
3805.00	3819.00	5075.00	5106.00	614.00	593.00	-----
1674.00	1645.00	6909.00	6771.00	2570.00	2518.00	-----
2228.00	2170.00	1911.00	1874.00	1997.00	1995.00	-----
3051.00	2974.00	3635.00	3584.00	6199.00	6176.00	-----
14433.00	14779.00	3975.00	3996.00	4913.00	5009.00	-----
1508.00	1475.00	897.00	864.00	2193.00	2184.00	-----
1865.00	1874.00	1410.00	1399.00	12585.00	12586.00	-----
2228.00	2269.00	5716.00	5790.00	1819.00	1821.00	-----
2055.00	2043.00	3306.00	3304.00	709.00	690.00	-----
1347.00	1325.00	3975.00	3893.00	6528.00	6484.00	-----
4492.00	4449.00	897.00	876.00	1944.00	1934.00	-----
3805.00	3757.00	7648.00	7475.00	3127.00	3138.00	-----
3051.00	3031.00	3635.00	3611.00	16492.00	16238.00	-----
1674.00	1676.00	838.00	824.00	2842.00	2741.00	-----
10030.00	10132.00	3053.00	2987.00	6528.00	6572.00	-----

4492.00	4478.00	10778.00	10756.00	10111.00	9669.00
3278.00	3209.00	1250.00	1234.00	11366.00	11758.00
1347.00	1339.00	2135.00	2108.00	4059.00	3979.00
8691.00	8644.00	5075.00	4993.00	1997.00	1990.00
1347.00	1331.00	897.00	875.00	2570.00	2445.00
4350.00	4341.00	1103.00	1082.00	2570.00	2576.00
12901.00	12601.00	3635.00	3602.00	11366.00	11207.00
3278.00	3339.00	11685.00	11548.00	10111.00	10101.00
12901.00	13224.00	1410.00	1385.00	13982.00	13668.00
3051.00	3072.00	2819.00	2826.00	-----	-----
5487.00	5453.00	19500.00	19348.00	-----	-----
2076.00	2128.00	4140.00	4057.00	-----	-----
1535.00	1554.00	2745.00	2716.00	-----	-----
3051.00	3063.00	1734.00	1710.00	-----	-----
4775.00	4741.00	3053.00	3104.00	-----	-----
10030.00	10021.00	8954.00	8789.00	-----	-----
-----	-----	1147.00	1151.00	-----	-----
-----	-----	2285.00	2234.00	-----	-----
-----	-----	897.00	894.00	-----	-----
-----	-----	688.00	677.00	-----	-----
-----	-----	966.00	955.00	-----	-----
-----	-----	688.00	669.00	-----	-----
FEATURES					
Source					
Location/Qualifiers					
1. .169234					
/organism="Homo sapiens"					
/db_xref="taxon:9606"					
/chromosome="7"					
/map="7p14-15"					
/clone="djs1 (RP11-16G1)"					
/cell_line="Male Blood"					
/clone_lib="RPC-11 Human Male BAC Library"					
392.494					
/rpt_family="MER21"					
complement(1075..1364)					
/rpt_family="Alu"					
complement(2210..2373)					
/standard_name="SWSS2626"					
/note="GenBank Accession Number: G12923"					
2594..2876					
/rpt_family="Alu"					
complement(3252..3702)					
/rpt_family="MLT1"					
complement(4489..4790)					
/rpt_family="Alu"					
complement(5307..5594)					
/rpt_family="Alu"					
5847..6152					
/rpt_family="Alu"					
complement(6970..7082)					
/standard_name="SWSS1000"					
/note="GenBank Accession Number: G00158"					
STS					
repeat_region					
repeat_region					
STS					
repeat_region					
repeat_region					
repeat_region					
repeat_region					
repeat_region					
STS					

```

repeat_region complement(7545..7634)
repeat_region /rpt_family="Alu"
repeat_region complement(8959..9186)
repeat_region /rpt_family="Alu"
repeat_region 10137..10423
repeat_region /rpt_family="Alu"

Query Match 2.6% Score 74; DB 9; Length 169234;
Best Local Similarity 100.0%; Pred. No. 3.7e-27; Mismatches 0; Indels 0; Gaps 0;
Matches 74; Conservative 0;

QY 1982 GGCACCTGTAATCCCGAGTCTTGGAGGCTGAGGAGGAGATCGCTTGAACCCAGGAG 2041
      |||||||
Db 31935 GGCACCTGTAATCCCGAGTCTTGGAGGCTGAGGAGGAGATCGCTTGAACCCAGGAG 31876

QY 2042 GTGAGGTTGCAGT 2055
      |||||||
Db 31875 GTGAGGTTGCAGT 31862

RESULT 77
AC093899/c
LOCUS AC093899 172816 bp DNA linear PRI 12-JUN-2002
DEFINITION Homo sapiens BAC clone RP11-724016 from 2, complete sequence.
ACCESSION AC093899 AC068884
VERSION AC093899.3 GI:18497265
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 172816)
Sulston, J.E. and Waterston, R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
99063792
9847074
2 (bases 1 to 172816)
Pearman, C., Haekenson, W. and Boyer, E.
The sequence of Homo sapiens BAC clone RP11-724016
Unpublished (2001)
3 (bases 1 to 172816)
Waterston, R.H.
Direct Submission
Submitted (10-SEP-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 172816)
Waterston, R.H.
Direct Submission
Submitted (05-FEB-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 172816)
Waterston, R.
Direct Submission
Submitted (12-JUN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Feb 5, 2002 this sequence version replaced gi:15625013.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
----- Summary Statistics
Center project name: H_NH0724016
Drafting Center: WIBR
-----

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality > 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:  
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Moon, P. Y., Zhao, B., Frengen, E., Tatenio, M., Catanese, J. J., and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>  
VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:  
Actual start of this clone is at base position 1 of RP11-724016; actual end is at base position 172816 of RP11-724016.

Unresolved simple sequence repeat from base position 22700 to 25900.

The sequence of AC068884 has been incorporated into AC093899.

FEATURES	Location/Qualifiers
source	1..172816
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="2"
	/map="2"
	/clone="RP11-724016"
	/clone_lib="RPCI-11"
repeat_region	19..1140
	/rpt_family="LI"
repeat_region	1141..1437
	/rpt_family="Alu"
repeat_region	1438..2657
	/rpt_family="LI"
repeat_region	2658..2729
	/rpt_family="LI"
repeat_region	2730..3320
	/rpt_family="LI"
repeat_region	3321..3642
	/rpt_family="Alu"
repeat_region	3643..4184
	/rpt_family="LI"
repeat_region	4185..4996
	/rpt_family="ERV1"
repeat_region	5011..5263
	/rpt_family="LI"
repeat_region	5285..5907
	/rpt_family="CRI"
repeat_region	6895..7078
	/rpt_family="LI"
repeat_region	7103..7421
	/rpt_family="Alu"
repeat_region	7807..7952
	/rpt_family="LI"
repeat_region	7981..8289
	/rpt_family="Alu"
repeat_region	8360..8551
	/rpt_family="LI"
repeat_region	8622..8718
	/rpt_family="MIR"

```
repeat_region 8784. .8825 /rpt_family="(TTTTC)n"
repeat_region 8840. .9134 /rpt_family="Alu"
repeat_region 9136. .9174 /rpt_family="Alu"
repeat_region 9861. .9881 /rpt_family="(A)n"
repeat_region 9882. .10166 /rpt_family="Alu"
repeat_region 16502. .16556 /rpt_family="L2"
repeat_region 16886. .17165 /rpt_family="L1"
repeat_region 17179. .17482 /rpt_family="Alu"
repeat_region 18025. .18119 /rpt_family="MIR"
repeat_region 18177. .18471 /rpt_family="Alu"
repeat_region 19312. .19441 /rpt_family="MER1_type"
repeat_region 19731. .20040 /rpt_family="MaLR"
repeat_region 20403. .20472 /rpt_family="L2"
repeat_region 21286. .21344 /rpt_family="(CA)n"
repeat_region 22724. .23977 /rpt_family="(TA)n"
repeat_region 24005. .24171 /rpt_family="(TA)n"
repeat_region 24196. .25007 /rpt_family="(TA)n"
repeat_region 25030. .25205 /rpt_family="(TA)n"
repeat_region 25223. .25398 /rpt_family="(TA)n"
repeat_region 25399. .25455 /rpt_family="AT_rich"
repeat_region 25456. .25919 /rpt_family="(TA)n"
repeat_region 25920. .26206 /rpt_family="Alu"
repeat_region 26893. .26917 /rpt_family="AT_rich"
repeat_region 27194. .27497 /rpt_family="L1"
repeat_region 27498. .27814 /rpt_family="MER2_type"
repeat_region 27815. .28151 /rpt_family="L1"
repeat_region 28152. .28278 /rpt_family="Alu"
repeat_region 28279. .28485 /rpt_family="L1"
repeat_region 28795. .28829 /rpt_family="AT_rich"
repeat_region 28986. .29100 /rpt_family="L2"
repeat_region 29237. .29349 /rpt_family="L1"
repeat_region 29350. .29484 /rpt_family="Alu"
repeat_region 29485. .29589 /rpt_family="L1"
repeat_region 29811. .29874 /rpt_family="L2"
repeat_region 29875. .30284 /rpt_family="L1"
repeat_region 30320. .30522 /rpt_family="L1"
repeat_region 30523. .30847 /rpt_family="L1"

/rpt_family="Alu"
30848. .30882
/rpt_family="L1"

Query Match 2.6%; Score 74; DB 9; Length 172816;
Best Local Similarity 100.0%; Pred. NO. 3.7e-27;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1982 GGCACCTGTAATCCCGAGCTACTTGGAGGCTCAGGCAGGAGAAATCGCTTGAACCCAGGAG 2041
|||||
Db 138970 GGCACCTGTAATCCCGAGCTACTTGGAGGCTCAGGCAGGAGAAATCGCTTGAACCCAGGAG 138911
|||||

QY 2042 GTGGAGGTTGCAGT 2055
|||||
Db 138910 GTGGAGGTTGCAGT 138897

RESULT 78
AC025518/c
LOCUS 176355 bp DNA linear PRI 29-JUN-2002
DEFINITION Homo sapiens chromosome 17, clone RP11-713H12, complete sequence.
ACCESSION AC025518
VERSION AC025518.8 GI:21629418
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 176355)
AUTHORS Birren,B., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 17, clone RP11-713H12
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 176355)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Bouckgalter,B., Brown,A., Burkett,G.,
Campiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (09-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 176355)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Bouckgalter,B., Brown,A., Camarata,J., Campiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collamore,A.,
Cooke,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K.,
Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N.,
Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
```





```
Query Match      2.6%; Score 74; DB 9; Length 176355;
Best Local Similarity 100.0%; Pred. No. 3.7e-27;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1982 GGCACCTATCCAGCTACTTGGGAGGCTGAGCGAGGAGAAATCGCTTGAACCCAGGAG 2041
DB 152206 GGCACCTATCCAGCTACTTGGGAGGCTGAGCGAGGAGAAATCGCTTGAACCCAGGAG 152147
QY 2042 GTGGAGGTTCAGT 2055
DB 152146 GTGGAGGTTCAGT 152133

RESULT 79
AC016772      176932 bp      DNA      linear      PRI 09-JAN-2002
LOCUS      Homo sapiens BAC clone RP11-567N19 from 2, complete sequence.
DEFINITION      AC016772
ACCESSION      AC016772.9 GI:18098289
VERSION      HTG.
KEYWORDS      Homo sapiens.
SOURCE      Homo sapiens.
ORGANISM      Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 176932)
AUTHORS      Sulston, J.E. and Waterston, R.
TITLE      Toward a complete human genome sequence
JOURNAL      Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE      99063792
PUBMED      9847074
REFERENCE      2 (bases 1 to 176932)
AUTHORS      Haglund, K., Abbott, A. and Spalding, L.
TITLE      The sequence of Homo sapiens BAC clone RP11-567N19
JOURNAL      Unpublished (2001)
REFERENCE      3 (bases 1 to 176932)
AUTHORS      Waterston, R.H.
TITLE      Direct Submission
JOURNAL      Submitted (04-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE      4 (bases 1 to 176932)
AUTHORS      Waterston, R.H.
TITLE      Direct Submission
JOURNAL      Submitted (19-APR-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE      5 (bases 1 to 176932)
AUTHORS      Waterston, R.H.
TITLE      Direct Submission
JOURNAL      Submitted (20-APR-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE      6 (bases 1 to 176932)
AUTHORS      Waterston, R.
TITLE      Direct Submission
JOURNAL      Submitted (07-NOV-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE      7 (bases 1 to 176932)
AUTHORS      Waterston, R.
TITLE      Direct Submission
JOURNAL      Submitted (09-JAN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT      On Jan 9, 2002 this sequence version replaced gi:13677173.
-----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
-----
Summary Statistics
-----
Center project name: H_NH0567N19
-----
```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACE3.6

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-401019; the clone sequenced to the right is AC068884. Actual end of this clone is at base position 176932 of RP11-567N19.

#### FEATURES

source

1..176932

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="2"

/map="2"

/clone="RP11-567N19"

/clone\_lib="RPCI-11"

1..303

/rpt\_family="MaLR"

239..639

/note="similar to Mus musculus EST AW988274 (NID:g8183383)"

305..811

/note="match to EST AA769615 (NID:g2820853) ob19g06.s1"

312..507

/note="match to EST AA918024 (NID:g3057914) ol73d08.s1"

379..1075

/note="similar to Mus musculus EST BF138850 (NID:g10977890)"

404..871

/note="similar to Mus musculus EST AA538008 (NID:g2284001)"

vj34h03.rl"

541..921

/note="similar to Mus musculus EST AA106541 (NID:gl657130)"

ml90d04.rl"

566..1074

/note="similar to Mus musculus EST B1455611 (NID:g15246267)"

703..1074

/note="match to EST B1550291 (NID:g15437603)"

747..1074

/note="similar to Bos taurus EST AV596137 (NID:g9713168)"

763..1076

/note="similar to Homo sapiens EST B1459746 (NID:g15250402)"

998..1074

/note="similar to EST B1976350 (NID:gl6350755)"

1550..1952

/rpt\_family="L1"

repeat\_region

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

repeat_region	2107..3018	Matches	74: Conservative	0: Mismatches	0: Indels	0: Gaps	0:
repeat_region	/rpt_family="L1"	1982	GGCACCTGTATCCAGCTACTTGGAGGCTTGAGCAGGAGCAATCGCTTGAACCCAGGAG	2041			
misc_feature	3425..3655						
misc_feature	3714..3942						
misc_feature	/note="match to EST AI801223 (NID:g5366695) to86d01.x1"						
misc_feature	3720..4209						
misc_feature	/note="match to EST AI424984 (NID:g4270915) tg38h07.x1"						
misc_feature	3720..3833						
misc_feature	/note="match to EST AA774547 (NID:g2833881) ai27a12.s1"						
misc_feature	4023..4155						
misc_feature	/note="similar to EST BI976350 (NID:gl6350755)"						
misc_feature	4035..4127						
misc_feature	/note="match to EST BG106656 (NID:gl12600502)"						
misc_feature	4038..4127						
misc_feature	/note="match to EST AA774547 (NID:g2833881) ai27a12.s1"						
misc_feature	4089..4101						
misc_feature	/note="similar to Homo sapiens EST BI459746 (NID:g15250402)"						
misc_feature	4274..4360						
misc_feature	/note="match to EST AV648795 (NID:g9869809)"						
misc_feature	4417..4659						
misc_feature	/note="match to EST AV648795 (NID:g9869809)"						
misc_feature	4489..4659						
misc_feature	/note="match to EST AA774547 (NID:g2833881) ai27a12.s1"						
misc_feature	4489..4659						
misc_feature	/note="match to EST BG106656 (NID:gl12600502)"						
misc_feature	4489..4659						
misc_feature	/note="match to EST BI550291 (NID:gl15437603)"						
misc_feature	4489..4659						
misc_feature	/note="similar to Bos taurus EST AV596137 (NID:g9713168)"						
misc_feature	4489..4659						
misc_feature	/note="similar to EST BI976350 (NID:gl6350755)"						
misc_feature	4489..4659						
misc_feature	/note="similar to Mus musculus EST BI455611 (NID:g15246267)"						
misc_feature	4610..4629						
misc_feature	/note="similar to Homo sapiens EST BI459746 (NID:g15250402)"						
repeat_region	4707..4999						
repeat_region	/rpt_family="Alu"						
repeat_region	4987..5025						
repeat_region	/rpt_family="AT-rich"						
repeat_region	5978..6366						
repeat_region	/rpt_family="ERV1"						
repeat_region	6370..6664						
misc_feature	/rpt_family="Alu"						
misc_feature	6773..6807						
misc_feature	/note="similar to Mus musculus EST BI455611 (NID:g15246267)"						
misc_feature	6773..6806						
misc_feature	/note="match to EST AA774547 (NID:g2833881) ai27a12.s1"						
misc_feature	6773..6806						
misc_feature	/note="similar to Bos taurus EST AV596137 (NID:g9713168)"						
misc_feature	6773..6805						
misc_feature	/note="match to EST AV648795 (NID:g9869809)"						
misc_feature	6773..6805						
misc_feature	/note="match to EST BG106656 (NID:gl12600502)"						
misc_feature	6773..6805						
misc_feature	/note="match to EST BI550291 (NID:gl15437603)"						
misc_feature	6773..6805						
misc_feature	/note="similar to EST BI976350 (NID:gl6350755)"						
misc_feature	7279..7555						
misc_feature	/note="match to EST BI550291 (NID:gl15437603)"						
misc_feature	7279..7454						
misc_feature	/note="similar to EST BI976350 (NID:gl6350755)"						
misc_feature	7313..7555						
misc_feature	/note="match to EST BI552902 (NID:gl15440214)"						
misc_feature	7398..7532						
misc_feature	/note="match to EST BG106656 (NID:gl12600502)"						
Query Match	2.6%; Score 74; DB 9; Length 176932;						
Best Local Similarity	100.0%; Pred. No. 3.7e-27;						
RESULT 80							
AC026130/c							
LOCUS	182909 bp DNA linear PRI 14-JUL-2002						
DEFINITION	Homo sapiens chromosome 17, clone RP11-801J18, complete sequence.						
ACCESSION	AC026130						
VERSION	AC026130.15 GI:217006699						
KEYWORDS	HTG.						
SOURCE	human.						
ORGANISM	Homo sapiens						
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.						
AUTHORS	1 (bases 1 to 182909)						
TITLE	Homo sapiens chromosome 17, clone RP11-801J18						
JOURNAL	Unpublished						
REFERENCE	2 (bases 1 to 182909)						
AUTHORS	Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F., Boguslavsky,L., Boukhgaiter,B., Brown,A., Burkett,G., Campyano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,I., Karitas,A., Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoucky,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPherson,R., Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenka,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pietrie,N., Pisan,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Tallmadge,J., Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trullio,J., Vassiliev,H., Viel,R., Vo.A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.						
TITLE	Direct Submission						
JOURNAL	Submitted (19-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA						
REFERENCE	3 (bases 1 to 182909)						
AUTHORS	Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgaiter,B., Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collamore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fato,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Gord,S., Graham,L., Grand-pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kellis,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenka,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rostov,P., Roman,J., Roy,A., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Tallmadge,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Viel,R., Vo.A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.						
TITLE	Direct Submission						
JOURNAL	Submitted (06-JUL-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA						

REFERENCE  
AUTHORS

4 (bases 1 to 182909)  
 Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,  
 Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,  
 Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collamore,A.,  
 Cook,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
 Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,  
 Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,  
 Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,  
 Karatas,A., Kellis,C., Landers,T., Levine,R., Lindblad-Toh,K.,  
 Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,  
 McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,  
 Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,  
 O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,  
 Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,  
 Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,  
 Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,  
 Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,  
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,  
 Zembek,L., Zimmer,A. and Zody,M.

TITLE  
JOURNAL

Submitted (14-JUL-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jul 6, 2002 this sequence version replaced gi:21426269.

## COMMENT

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WTBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence.submissions@genome.wi.mit.edu](mailto:sequence.submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L5701

Center clone name: 801\_J\_18

FEATURES  
source

----- Location/Qualifiers  
 1..182909  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="17"  
 /map="17"  
 /clone="RP11-801J18"  
 /clone\_lib="RPC1-11 Human Male BAC"  
 93..407  
 /rpt\_family="AluSg"  
 complement(1561..1784)  
 /rpt\_family="L1MB7"  
 complement(1789..2051)  
 /rpt\_family="AluSg"  
 complement(4996..5042)  
 /rpt\_family="L2"  
 6912..7047  
 /rpt\_family="FLAM\_C"  
 9254..9325  
 /rpt\_family="L2"  
 complement(11585..11631)  
 /rpt\_family="L2"  
 12512..12811  
 /rpt\_family="AluSg"  
 complement(13375..13484)  
 /rpt\_family="L2"  
 complement(13603..13706)  
 /rpt\_family="MIR"  
 complement(14123..14443)  
 /rpt\_family="AluSx"  
 14533..14990  
 /rpt\_family="L2"  
 complement(15253..15339)  
 /rpt\_family="L2"  
 16201..16279  
 /rpt\_family="MIR"  
 complement(18078..18364)  
 /rpt\_family="AluY"

repeat\_region  
 complement(18391..18472)  
 /rpt\_family="L1MB4A"  
 18529..18582  
 /rpt\_family="(TATATG)n"  
 complement(19092..19180)  
 /rpt\_family="L1MB3"  
 20169..20453  
 /rpt\_family="AluJb"  
 20454..20497  
 /rpt\_family="(CA)n"  
 20552..20576  
 /rpt\_family="(CA)n"  
 complement(20585..20813)  
 /rpt\_family="L2"  
 complement(20874..21062)  
 /rpt\_family="L2"  
 complement(21389..21395)  
 /note="<30 qual SNGL region"  
 complement(21443..21745)  
 /rpt\_family="AluY"  
 complement(21508..21516)  
 /note="<30 qual SNGL region"  
 complement(21530..21538)  
 /note="<30 qual SNGL region"  
 22759..23068  
 /rpt\_family="AluSx"  
 23122..23128  
 /note="<30 qual SNGL region"  
 23138..23151  
 /note="<30 qual SNGL region"  
 complement(23462..23640)  
 /rpt\_family="L2"  
 26691..26909  
 /rpt\_family="L1MC5"  
 complement(27053..27364)  
 /rpt\_family="AluSg"  
 complement(27417..27477)  
 /rpt\_family="U7"  
 complement(27487..27525)  
 /rpt\_family="L1MA2"  
 27526..27586  
 /rpt\_family="(TA)n"  
 complement(27587..27659)  
 /rpt\_family="L1MA2"  
 complement(28085..28239)  
 /rpt\_family="L1MB3"  
 28304..28561  
 /rpt\_family="AluY"  
 28831..28851  
 /rpt\_family="AT\_Rich"  
 complement(29182..29789)  
 /rpt\_family="MER7C"  
 30200..30435  
 /rpt\_family="AluJo"  
 30436..30555  
 /rpt\_family="AluJo/FRAM"  
 complement(30603..30912)  
 /rpt\_family="AluSg"  
 complement(32603..32903)  
 /rpt\_family="AluSx"  
 33422..33541  
 /rpt\_family="MIR"  
 complement(35289..35601)  
 /rpt\_family="AluSp"  
 36008..36054  
 /rpt\_family="L3"  
 36097..36121

## Query Match

Best Local Similarity 100.0%; Score 74; DB 9; Length 182909;

Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1982 GGCACCTGTATCCAGCTACTTGGAGGCTGAGCAGAGATCGCTTGAACCCAGCAG 2041

```
|||||
88443 GGCACCTGTAATCCAGCTACTTGGAGGCTGAGCAGGAGATCGCTTGAACCCAGGAG 88384
QY 2042 GTGGAGGTTGCAGT 2055
Db 88383 GTGGAGGTTGCAGT 88370

RESULT 81
AC116170 186870 bp DNA linear HTG 24-AUG-2002
LOCUS Homo sapiens chromosome 17 clone RP11-855E10 map 17, WORKING DRAFT
DEFINITION SEQUENCE, 4 unordered pieces.
AC116170
VERSION AC116170.2 GI:22474934
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 186870)
Birren,B., Nussbaum,C. and Lander,E.
Homo sapiens chromosome 17, clone RP11-855E10
Unpublished
2 (bases 1 to 186870)
Birren,B., Linton,L., Nussbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,
Boukhaltier,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
Lander,T., Lehoczyk,J., Levine,R., Lindblad-Toh,K., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Meneus,L.,
Mihoval,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (25-MAR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 186870)
Birren,B., Nussbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhaltier,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,K., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Macdonald,P., Major,J., Matthews,C.,
Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrum,J., Meneus,L., Mihoval,T., Mienga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schuback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 24, 2002 this sequence version replaced gi:19703298.
```

```
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L26190
Center clone name: 855_E10
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 18544 bases at least Q40
Consensus quality: 185963 bases at least Q30
Consensus quality: 186196 bases at least Q20
Insert size: 188000; agarose-fp
Quality coverage: 17.0 in Q20 bases; agarose-fp
Quality coverage: 17.1 in Q20 bases; sum-of-continigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 16661: contig of 16661 bp in length
* 16662 16761: gap of 100 bp
* 16762 42342: contig of 25581 bp in length
* 42343 42442: gap of 100 bp
* 42443 114057: contig of 71615 bp in length
* 114058 114157: gap of 100 bp
* 114158 186870: contig of 72713 bp in length.
*
* Location/Qualifiers
* source
* 1-186870
* /organism="Homo sapiens"
* /db_xref="taxon:9606"
* /chromosome="17"
* /map="17"
* /clone.lib="RP11-855E10"
* /clone.lib="RP11-855E10"
* /clone.lib="RP11-855E10"
* 1-16661
* /note="assembly_fragment"
* clone_end:SP6
* vector_side:left
* 16762..42342
* /note="assembly_fragment"
* 42443..114057
* /note="assembly_fragment"
* 114158..186870
* /note="assembly_fragment"
* 300 others
BASE COUNT 51573 a 40568 c 39916 g 54513 t
ORIGIN
Query Match 2.6%; Score 74; DB 2; Length 186870;
Best Local Similarity 100.0%; Pred.No. 3.7e-27;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1982 GGCACCTGTAATCCAGCTACTTGGAGGCTGAGCAGGAGATCGCTTGAACCCAGGAG 2041
Db 138977 GGCACCTGTAATCCAGCTACTTGGAGGCTGAGCAGGAGATCGCTTGAACCCAGGAG 139016
QY 2042 GTGGAGGTTGCAGT 2055
Db 139037 GTGGAGGTTGCAGT 139050
RESULT 82
```

AC063929 187697 bp DNA linear HTG 09-MAY-2002  
LOCUS Homo sapiens chromosome 3 clone RP11-64L3, WORKING DRAFT SEQUENCE,  
DEFINITION 13 unordered pieces.  
AC063929  
VERSION AC063929.15 GI:20335598  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 187697)  
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayale,M., Banks,T.,  
Barbata,J., Benton,J., Blinag,K., Blankenburg,K., Bonnin,D.,  
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,  
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,  
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,  
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,  
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,  
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,  
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,  
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,  
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,  
Gorelli,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,  
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,  
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,  
Homi,F., Howard,S., Huber,J., Huly,S., Hume,J., Jackson,L.E.,  
Jacobson,B., Jia,Y., Johnson,R., Jollivet,S., Joudah,S.,  
Karlisson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,  
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,  
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,  
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,  
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,  
Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,  
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,  
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,  
Nguyen,N., Nickerson,E., Nwokkwo,S., Ogih,M., Okwuonu,G.,  
Oraquyne,N., Oviedo,E., Pace,A., Payton,B., Peary,J., Perez,L.,  
Peters,L., Pickens,R., Primus,E., Pui,L., Quiles,M., Ren,Y.,  
Rives,M., Rojas,A., Rojibokan,I., Roife,M., Ruiz,S., Savery,G.,  
Scherer,S., Scott,G., Shen,H., Shoostrari,N., Sisson,I.,  
Sodergren,E., Sonaite,T., Sparks,A., Stanley,H., Stone,H.,  
Sutton,A., Svatek,A., Tabor,P., Tamerisa,K., Tamerisa,K., Tang,H.,  
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,  
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,  
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,  
Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K.,  
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
Weinstock,G. and Gibbs,R.  
Direct Submission  
Unpublished  
2 (bases 1 to 187697)  
Worley,K.C.  
Direct Submission  
Submitted (22-APR-2000) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 187697)  
Worley,K.C.  
Direct Submission  
Submitted (09-MAY-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Apr 28, 2002 this sequence version replaced gi:16117928.  
----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: http://www.hgsc.bcm.tmc.edu/  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
-----

Center project name: HAYV  
Center clone name: RP11-64L3  
----- Summary Statistics  
Sequencing vector: M13;  
Chemistry: Dye-primer Bodipy: 54% of reads  
Chemistry: Dye-terminator Big Dye: 46% of reads  
Assembly program: Phrap: version 0.990329  
Consensus quality: 178595 bases at least Q40  
Consensus quality: 182392 bases at least Q30  
Consensus quality: 184739 bases at least Q20  
Estimated insert size: 184304; sum-of-contigs estimation  
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank\_draft\_data.html).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 13 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1  
\* 4028: contig of 4028 bp in length  
\* 4128: gap of unknown length  
\* 4129: 7656: contig of 3528 bp in length  
\* 7657: 7756: gap of unknown length  
\* 7757: 11625: contig of 3869 bp in length  
\* 11626: 11725: gap of unknown length  
\* 11726: 17061: contig of 5336 bp in length  
\* 17062: 17161: gap of unknown length  
\* 24677: 24676: contig of 7515 bp in length  
\* 24677: 24776: gap of unknown length  
\* 24777: 35587: contig of 10811 bp in length  
\* 35588: 35887: gap of unknown length  
\* 35888: 48587: contig of 12990 bp in length  
\* 48588: 48587: gap of unknown length  
\* 61359: 61358: contig of 12671 bp in length  
\* 61359: 61459: gap of unknown length  
\* 61459: 76906: contig of 15448 bp in length  
\* 76907: 77006: gap of unknown length  
\* 77007: 97133: contig of 20127 bp in length  
\* 97134: 97233: gap of unknown length  
\* 97234: 120991: contig of 23758 bp in length  
\* 120992: 121091: gap of unknown length  
\* 121092: 152130: contig of 31039 bp in length  
\* 152131: 152230: gap of unknown length  
\* 152231: 187697: contig of 35467 bp in length.  
FEATURES  
Location/Qualifiers  
source  
1..187697  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="3"  
/clone="RP11-64L3"  
BASE COUNT 53460 a 40236 c 39647 g 53120 t 1234 others  
ORIGIN  
Query Match 2.6%; Score 74; DB 2; Length 187697;  
Best Local Similarity 100.0%; Pred. No. 3.7e-27;  
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1982 GGCACCTGTAAATCCCGAGCTACTTGGAGGCTGAGCAGGAGATCGCTTGAACCCAGGAG 2041  
|||||  
Db 443 GGCACCTGTAAATCCCGAGCTACTTGGAGGCTGAGCAGGAGATCGCTTGAACCCAGGAG 502  
QY 2042 GTGGAGGTTGCAGT 2055  
|||||  
Db 503 GTGGAGGTTGCAGT 516  
RESULT 83  
AC092700 201886 bp DNA linear PRI 06-JUN-2002  
LOCUS

DEFINITION  
AC092700  
AC092700.2  
KEYWORDS  
SOURCE

Homo sapiens chromosome 8, clone RP11-661A3, complete sequence.  
AC092700  
AC092700.2 GI:21327455  
HTG  
human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 201886)

AUTHORS

Birren, B., Linton, L., Nusbaum, C. and Lander, E.

TITLE

Homo sapiens chromosome 8, clone RP11-661A3

REFERENCE

2 (bases 1 to 201886)

AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,  
Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhvalter, B.,  
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,  
Chospel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,  
Cooke, P., DeArelano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S.,  
Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,  
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,  
Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K.,  
Lamazares, R., Landers, T., Lehoczký, J., Levine, R., Liu, G.,  
Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,  
McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J.,  
Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C.,  
Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,  
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,  
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,  
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R.,  
Seaman, S., Severy, P., Sougne, C., Spencer, B., Stange-Thomann, N.,  
Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S.,  
Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J.,  
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,  
Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

TITLE

Submitted (20-JUL-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA

JOURNAL

Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE

3 (bases 1 to 201886)

AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,  
Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L.,  
Boukhvalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,  
Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,  
Cook, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J. S., Dodge, S.,  
Faro, S., Ferreira, P., FitzGerald, M., FitzHugh, W., Gage, D.,  
Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L.,  
Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I.,  
Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K.,  
Lamazares, R., Landers, T., Lehoczký, J., Levine, R., Lindblad-Toh, K.,  
Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N.,  
Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrim, J.,  
Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C.,  
Nicol, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P.,  
O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,  
Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C.,  
Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S.,  
Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N.,  
Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S.,  
Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J.,  
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,  
Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

TITLE

Submitted (06-JUN-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA

JOURNAL

Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT

On Jun 6, 2002 this sequence version replaced gi:14971446.

All repeats were identified using RepeatMasker:

Smith, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information  
Center project name: L11751  
Center clone name: 661\_A3  
-----

FEATURES

source

Location/Qualifiers

1. .201886  
/organism:"Homo sapiens"  
/db\_xref:"taxon:9606"  
/chromosome:"8"  
/map:"8"  
/clone:"RP11-661A3"  
/clone\_lib:"RPC1-11 Human Male BAC"  
4356. .4523  
/rpt\_family:"HAL1b"  
complement(4625. .4906)  
/rpt\_family:"AluJb"  
8941. .9014  
/rpt\_family:"AluJ/FRAM"  
9122. .9238  
/rpt\_family:"(TA)n"  
11059. .11102  
/rpt\_family:"L3"  
complement(11461. .11765)  
/rpt\_family:"AluSx"  
12218. .12291  
/rpt\_family:"tRNA-Glu-GAG"  
12584. .12885  
/rpt\_family:"AluSx"  
16513. .16567  
/rpt\_family:"L2"  
16626. .16926  
/rpt\_family:"AluY"  
18349. .18494  
/rpt\_family:"AluSg/x"  
18705. .19009  
/rpt\_family:"AluSx"  
complement(20040. .20229)  
/rpt\_family:"MIR"  
20313. .20612  
/rpt\_family:"AluSx"  
20813. .21145  
/rpt\_family:"LIME4A"  
22622. .22652  
/rpt\_family:"AT-rich"  
22708. .22812  
/rpt\_family:"LIME4A"  
24360. .24382  
/rpt\_family:"AT-rich"  
24482. .24545  
/rpt\_family:"(TA)n"  
complement(24715. .25019)  
/rpt\_family:"AluJb"  
25673. .25879  
/rpt\_family:"AluSg/x"  
26165. .26219  
/rpt\_family:"(TA)n"  
complement(26405. .26553)  
/rpt\_family:"AluJb"  
26671. .26899  
/rpt\_family:"MIR"  
27947. .28028  
/rpt\_family:"MIR"  
30867. .31239  
/rpt\_family:"LIMCC"  
31240. .31487  
/rpt\_family:"AluJb"  
31488. .31862  
/rpt\_family:"LIMCC"  
31923. .32003  
/rpt\_family:"MER21B"  
32024. .32771  
/rpt\_family:"MER21B"  
32826. .33568

repeat\_region



COMMENT

Baylor Plaza, Houston, TX 77030, USA  
On Sep 30, 2001 this sequence version replaced gi:15799479.  
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email  
[gc-help@bcm.tmc.edu](mailto:gc-help@bcm.tmc.edu)

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.  
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.  
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT.

FEATURES  
Source

1..205952  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="12"  
/clone="RP11-627K11"

misc\_feature

1..2005  
/note="overlaps bases 104273..106277 of clone AC068792"  
/function="clone overlap"

repeat\_region

14..206  
/rpt\_family="MIR"

repeat\_region

complement(1105..1412)  
/rpt\_family="AluSx"

repeat\_region

complement(1413..1679)  
/rpt\_family="LTR48"

repeat\_region

1706..1987  
/rpt\_family="AluSx"

repeat\_region

1988..2015  
/rpt\_family="AluSx"

repeat\_region

2022..2188  
/rpt\_family="AluSx"

repeat\_region

2189..2400  
/rpt\_family="AluSg/x"

repeat\_region

complement(2422..2724)  
/rpt\_family="AluSg"

repeat\_region

complement(2737..3047)  
/rpt\_family="AluSx"

repeat\_region

3112..3395  
/rpt\_family="MLT2E"

repeat\_region

3481..3516  
/rpt\_family="MLT1A1"

repeat\_region

complement(3517..3760)  
/rpt\_family="AluSx"

repeat\_region 4039..4125  
/rpt\_family="AluJo"  
repeat\_region 4130..4433  
/rpt\_family="AluSg"  
repeat\_region 4695..5130  
/rpt\_family="L2"  
repeat\_region complement(5131..5432)  
/rpt\_family="AluSg"  
repeat\_region 5433..5454  
/rpt\_family="L2"  
repeat\_region complement(5455..5748)  
/rpt\_family="AluSg"  
repeat\_region 5749..5822  
/rpt\_family="L2"  
repeat\_region complement(5833..6112)  
/rpt\_family="AluJo"  
repeat\_region 6169..6207  
/rpt\_family="AT-rich"  
repeat\_region complement(6416..6702)  
/rpt\_family="AluSx"  
repeat\_region 7151..7185  
/rpt\_family="A-rich"  
repeat\_region 7220..7540  
/rpt\_family="AluJb"  
repeat\_region complement(7649..7952)  
/rpt\_family="AluSx"  
repeat\_region complement(8162..8475)  
/rpt\_family="AluY"  
repeat\_region complement(8477..8768)  
/rpt\_family="AluJb"  
repeat\_region complement(8838..8993)  
/rpt\_family="MERSA"  
STS 9177..9471  
/standard\_name="103578"  
repeat\_region 9206..9338  
/rpt\_family="MIR"  
repeat\_region complement(9422..9506)  
/rpt\_family="L2"  
repeat\_region complement(9560..9858)  
/rpt\_family="AluSg1"  
repeat\_region complement(10056..10174)  
/rpt\_family="FLAM\_C"  
repeat\_region 10225..10523  
/rpt\_family="AluSg"  
repeat\_region 10568..10850  
/rpt\_family="AluSg"  
repeat\_region 10851..10873  
/rpt\_family="CAAA)n"  
repeat\_region 11010..11059  
/rpt\_family="TA)n"

Query Match 2.6% Score 74; DB 9; Length 205952;

Best Local Similarity 100.0%; Pred. No. 3.7e-27;  
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1982 GGCACCTCTAATCCAGCTACTTGGAGGCTTGAGCAGAGAGATCGCTTGAACTCAGGAG 2041  
|||||  
Db 33550 GGCACCTCTAATCCAGCTACTTGGAGGCTTGAGCAGAGAGATCGCTTGAACTCAGGAG 33609  
|||||

QY 2042 GTGGAGGTTGCAGT 2055  
|||||

Db 33610 GTGGAGGTTGCAGT 33623  
|||||

RESULT 85

CNS01DWD/c

LOCUS

DEFINITION

Human chromosome 14 DNA sequence

from chromosome 14 of Homo sapiens (human), complete sequence.

ACCESSION

AL137128

VERSION

AL137128.4

KEYWORDS

HTG.

SOURCE

228652 bp DNA linear  
PRI 40-APR-2001  
HAC R-951P22 of library R951-11



ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 228652)  
 AUTHORS Heilig, R., Petit, J. L., Vico, V., Dasilva, C., Robert, C., Wincker, P.,  
 Brothier, P., Cattolico, L., Barbe, V., Pelletier, E., Artiguenave, F.,  
 Levy, M., Eckenberg, R., Bruls, F., Deberardinis, V., Cruaud, C.,  
 Gypay, G., Saurin, W. and Weissensbach, J.  
 TITLE Sequencing of the human chromosome 14  
 REFERENCE Unpublished  
 2 (bases 1 to 228652)  
 AUTHORS Genoscope.  
 TITLE Direct Submission  
 JOURNAL Submitted (30-APR-2001) Genoscope - Centre National de Sequençage ;  
 BP 111 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
 - web : www.genoscope.cns.fr)  
 COMMENT On May 2, 2001 this sequence version replaced gi:9755791.  
 ----- Genome Center  
 Center: Genoscope / Centre National de Sequençage  
 Center code: GS  
 Web site: http://www.genoscope.cns.fr/  
 Contact: seqref@genoscope.cns.fr

```

repeat_region 4263. .4293
/rpt_family="not_experimental
/rpt_family="(TTTA)n"
complement(4294. .4574)
/evidence=not_experimental
/rpt_family="AluY"
5229. .5498
/evidence=not_experimental
/rpt_family="AluJo"
complement(5615. .5870)
/evidence=not_experimental
/rpt_family="AluJb"
complement(5962. .6152)
/evidence=not_experimental
/rpt_family="MER3"
complement(6266. .6560)
/evidence=not_experimental
/rpt_family="AluJb"
complement(8253. .8442)
/evidence=not_experimental
/rpt_family="MER58A"
complement(8754. .8997)
/evidence=not_experimental
/rpt_family="MER20"
9183. .9210
/evidence=not_experimental
/rpt_family="AT_rich"
9428. .9463
/evidence=not_experimental
/rpt_family="AT_rich"
complement(10944. .11037)
/evidence=not_experimental
/rpt_family="MER58"
complement(11173. .11405)
/evidence=not_experimental
/rpt_family="MER46A"
11785. .11866
/evidence=not_experimental
/rpt_family="L1MC5"
11882. .12020
/evidence=not_experimental
/rpt_family="MER63"
12117. .12151
/evidence=not_experimental
/rpt_family="AT_rich"
12155. .12451
/evidence=not_experimental
/rpt_family="AluJo"
13344. .13572
/evidence=not_experimental
/rpt_family="MIR"
14570. .14633
/evidence=not_experimental
/rpt_family="AT_rich"
complement(15123. .15591)
/evidence=not_experimental
/rpt_family="L1MC3e"
18204. .18500
/evidence=not_experimental
/rpt_family="AluJo"
18840. .18862
/evidence=not_experimental
/rpt_family="AT_rich"
18893. .19050
/evidence=not_experimental
/rpt_family="L1ME2"
19086. .19110
/evidence=not_experimental
/rpt_family="(A)n"
complement(19754. .20039)
/evidence=not_experimental
/rpt_family="AluJb"
22669. .22969

```

```

repeat_region /evidence=not_experimental
/rpt_family="AluSq"
23510. .23619
/evidence=not_experimental
/rpt_family="MIR"
23817. .24117
/evidence=not_experimental
/rpt_family="AluSq"
complement(24121. .24252)
/evidence=not_experimental
/rpt_family="MIR"
complement(24667. .24959)
/evidence=not_experimental
/rpt_family="AluSx"
complement(26086. .26158)
/evidence=not_experimental
/rpt_family="L2"
26183. .26302
/evidence=not_experimental
/rpt_family="MIR"
26303. .26613
/evidence=not_experimental
/rpt_family="AluJb"
26614. .26704
/evidence=not_experimental
/rpt_family="MIR"
26707. .26891
/evidence=not_experimental
/rpt_family="MIR"
complement(27533. .28017)
/evidence=not_experimental
/rpt_family="LTR35"
complement(28078. .28837)
/evidence=not_experimental
/rpt_family="LTR8A"
29031. .29346
/evidence=not_experimental
/rpt_family="AluJo"
29349. .29660
/evidence=not_experimental
/rpt_family="AluJb"
complement(29674. .29730)
/evidence=not_experimental
/rpt_family="L2"
29731. .29780
/evidence=not_experimental
/rpt_family="(CA)n"
complement(29781. .30315)
/evidence=not_experimental
/rpt_family="L2"
complement(31044. .31338)
/evidence=not_experimental
/rpt_family="AluSq"
31497. .31664
/evidence=not_experimental
/rpt_family="HAL1"
31665. .31944
/evidence=not_experimental
/rpt_family="AluSp"
31945. .31968
/evidence=not_experimental
/rpt_family="(CA)n"
31969. .32285
/evidence=not_experimental
/rpt_family="AluY"
32286. .32624
/evidence=not_experimental
/rpt_family="HAL1"

```

Query Match 2.6%; Score 74; DB 9: Length 250681;  
 Best Local Similarity 100.0%; Pred.No. 3.7e-27;  
 Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1986 CCTGTATCCAGCTACTTGGAGGCTGAGCAGAGAAATCGCTTGAACCCAGGAGTGG 2045  
Db 236194 CTTGTAATCCAGCTACTTGGAGGCTGAGCAGAGAAATCGCTTGAACCCAGGAGTGG 236253

QY 2046 AGGTGCAGTAAGC 2059  
Db 236254 AGGTGCAGTAAGC 236267

RESULT 87  
AL512274

LOCUS AL512274 79319 bp DNA linear PRI 03-JUL-2001  
DEFINITION Human DNA sequence from clone Rp11-7k24 on chromosome 6, complete  
sequence.

ACCESSION AL512274  
VERSION AL512274.9 GI:14596369  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 79319)  
Hall, R.  
Direct Submission  
Submitted (03-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
requests: clonerequest@sanger.ac.uk  
On Jul 4, 2001 this sequence version replaced gi:13751511.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. The following  
abbreviations are used to associate primary accession numbers given  
in the feature table with their source databases: Em, EMBL; SW:,  
SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP  
database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep/ This sequence  
was generated from part of bacterial clone contigs of human  
chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping  
Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr6  
Rp11-7k24 is from the library RPCI-11.1 constructed by the group of  
Pieter de Jong. For further details see  
http://www.chori.org/bacpac/home.htm  
VECTOR: pBACe3.6  
IMPORTANT: This sequence is not the entire insert of clone  
Rp11-7k24 it may be shorter because we sequence overlapping  
sections only once, except for a 100 base overlap.  
The true left end of clone Rp11-7k24 is at 1 in this sequence. The  
true left end of clone Rp1-139D8 is at 77320 in this sequence. The  
true right end of clone Rp11-533020 is at 40694 in this sequence.

FEATURES  
source  
1..79319  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/clone="Rp11-7k24"  
/clone\_lib="RPCI-11.1"  
7..387  
misc\_feature /note="match: GSS: Em:AQ135411"  
14..484  
misc\_feature /note="match: GSS: Em:B72192"  
18..461  
misc\_feature /note="match: GSS: Em:AQ440906"

repeat\_region 495..703  
/note="MER58A repeat: matches 2..223 of consensus"  
785..1093  
/note="AluX repeat: matches 1..306 of consensus"  
complement(1337..1795)  
/note="match: GSS: Em:AQ177716"  
1826..1959  
/note="FLAM\_C repeat: matches 1..131 of consensus"  
2290..2586  
/note="AluX repeat: matches 1..301 of consensus"  
2880..3136  
/note="MLTIC repeat: matches 191..464 of consensus"  
3148..3289  
/note="LTR8 repeat: matches 444..584 of consensus"  
3290..3592  
/note="AluY repeat: matches 1..303 of consensus"  
3593..4032  
/note="LTR8 repeat: matches 1..444 of consensus"  
complement(4005..4217)  
/note="match: GSS: Em:AQ209273"  
complement(4016..4252)  
/note="match: GSS: Em:AQ462798"  
4342..4400  
/note="MER4-internal repeat: matches 6115..6171 of  
consensus"  
4401..5356  
/note="LTR25-internal repeat: matches 5493..6446 of  
consensus"  
5357..5580  
/note="MER4-internal repeat: matches 4806..6115 of  
consensus"  
5563..6703  
/note="MER4-internal repeat: matches 2321..5179 of  
consensus"  
6718..7027  
/note="AluSc repeat: matches 1..303 of consensus"  
7106..7317  
/note="MLTIC repeat: matches 1..222 of consensus"  
7358..7597  
/note="AluJo repeat: matches 54..291 of consensus"  
7800..7952  
/note="FLAM\_C repeat: matches 3..143 of consensus"  
7953..8103  
/note="FRAM repeat: matches 5..156 of consensus"  
8444..8742  
/note="AluX repeat: matches 1..291 of consensus"  
8880..9166  
/note="AluY repeat: matches 16..303 of consensus"  
9252..9560  
/note="AluX repeat: matches 1..309 of consensus"  
10083..10388  
/note="AluSq repeat: matches 1..306 of consensus"  
11102..11153  
/note="L2 repeat: matches 2659..2710 of consensus"  
11170..11339  
/note="MIR repeat: matches 75..239 of consensus"  
11418..11702  
/note="AluX repeat: matches 1..289 of consensus"  
11758..12052  
/note="AluY repeat: matches 1..300 of consensus"  
complement(11844..12204)  
/note="match: GSS: Em:AQ253047"  
12346..12549  
/note="MIR repeat: matches 15..262 of consensus"  
13007..13103  
/note="L2 repeat: matches 2602..2707 of consensus"  
13150..14317  
/note="L2 repeat: matches 1238..2476 of consensus"  
14324..14394  
/note="MER3 repeat: matches 137..208 of consensus"  
14395..14691  
/note="AluY repeat: matches 1..293 of consensus"  
14692..14813



Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
 Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
 Kamat, A., Karatas, A., Kells, C., Laroque, K., Lamazares, R.,  
 Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G.,  
 MacLean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C.,  
 McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Meneus, L.,  
 Mhova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,  
 Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,  
 Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,  
 Raymond, C., Retta, R., Riback, M., Riley, R., Rise, C., Rogov, P.,  
 Roman, J., Rosetti, P., Roy, A., Santos, R., Schauer, S., Schuback, R.,  
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
 Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
 Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,  
 Viell, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,  
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (17-MAY-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On May 17, 2002 this sequence version replaced gi:20128778.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WBIR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L2348  
 Center clone name: 197\_f\_11  
 -----

Only the last 96.2 kilobases of this clone are being submitted.  
 The remainder overlaps accession number AC021915 [WICGR project  
 L4214].

FEATURES	source	Location/Qualifiers
		1. .96165 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="RP11-197111" /clone_lib="RP11-11 Human Male BAC" complement(1311..1698) /rpt_family="MLT1B" 1781..2123 /rpt_family="MLT1A1" complement(2926..3341) /rpt_family="LTRL6A1" complement(2926..3290) /rpt_family="LTRL6A" complement(3471..3684) /rpt_family="L2" complement(4149..4241) /rpt_family="MIR" 5031..5054 /rpt_family="AT_rich" 5456..5592 /rpt_family="MER91A" 5613..5700 /rpt_family="FLAM_A" 5876..6185 /rpt_family="AluSg" complement(6567..6698) /rpt_family="MER31-int" 6926..7041 /rpt_family="AluJb" 7042..7354 /rpt_family="AluSg" 7355..7547 /rpt_family="AluJb" complement(7903..7973) /rpt_family="MER92C" complement(8035..8210)
repeat_region		/rpt_family="MER2" complement(17447..17764)
repeat_region		/rpt_family="MER2" 17765..17974
repeat_region		/rpt_family="MER2" 19797..20025
repeat_region		/rpt_family="L2" 20478..20704
repeat_region		/rpt_family="L2" 20783..21118
repeat_region		/rpt_family="L1M4" 21819..22345
repeat_region		/rpt_family="MER77" 22873..23048
repeat_region		/rpt_family="MLT1H" 23295..23411
repeat_region		/rpt_family="MLT1H" 23543..23703
repeat_region		/rpt_family="MIR" 23943..24001
repeat_region		/rpt_family="(TA)n" 24125..24238
repeat_region		/rpt_family="ALUSX" 12508..12579
repeat_region		/rpt_family="MLT1K" 13584..13613
repeat_region		/rpt_family="AT_rich" 13845..13897
repeat_region		/rpt_family="(CA)n" 13939..13991
repeat_region		/rpt_family="(TGGA)n" 14001..14142
repeat_region		/rpt_family="MIR3" 15420..15449
repeat_region		/rpt_family="(TC)n" complement(15450..15615)
repeat_region		/rpt_family="AluJo" 15616..15722
repeat_region		/rpt_family="CT-rich" complement(15723..15849)
repeat_region		/rpt_family="AluJo" 16353..16387
repeat_region		/rpt_family="AT_rich" 16988..17165
repeat_region		/rpt_family="L1ME" 17292..17446
repeat_region		/rpt_family="MER2" complement(17447..17764)
repeat_region		/rpt_family="AluJb" 17765..17974
repeat_region		/rpt_family="MER2" 19797..20025
repeat_region		/rpt_family="L2" 20478..20704
repeat_region		/rpt_family="L2" 20783..21118
repeat_region		/rpt_family="L1M4" 21819..22345
repeat_region		/rpt_family="MER77" 22873..23048
repeat_region		/rpt_family="MLT1H" 23295..23411
repeat_region		/rpt_family="MLT1H" 23543..23703
repeat_region		/rpt_family="MIR" 23943..24001
repeat_region		/rpt_family="(TA)n" 24125..24238
repeat_region		/rpt_family="ALUSX" 12508..12579
repeat_region		/rpt_family="MLT1K" 13584..13613
repeat_region		/rpt_family="AT_rich" 13845..13897
repeat_region		/rpt_family="(CA)n" 13939..13991
repeat_region		/rpt_family="(TGGA)n" 14001..14142
repeat_region		/rpt_family="MIR3" 15420..15449
repeat_region		/rpt_family="(TC)n" complement(15450..15615)
repeat_region		/rpt_family="AluJo" 15616..15722
repeat_region		/rpt_family="CT-rich" complement(15723..15849)
repeat_region		/rpt_family="AluJo" 16353..16387
repeat_region		/rpt_family="AT_rich" 16988..17165
repeat_region		/rpt_family="L1ME" 17292..17446
repeat_region		/rpt_family="MER2" complement(17447..17764)
repeat_region		/rpt_family="AluJb" 17765..17974
repeat_region		/rpt_family="MER2" 19797..20025
repeat_region		/rpt_family="L2" 20478..20704
repeat_region		/rpt_family="L2" 20783..21118
repeat_region		/rpt_family="L1M4" 21819..22345
repeat_region		/rpt_family="MER77" 22873..23048
repeat_region		/rpt_family="MLT1H" 23295..23411
repeat_region		/rpt_family="MLT1H" 23543..23703
repeat_region		/rpt_family="MIR" 23943..24001
repeat_region		/rpt_family="(TA)n" 24125..24238

Query Match 2.6%; Score 73; DB 9; Length 96165;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-26;  
 Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1983	GCACCTGTATCCAGCTACTTGGGAGGCTGAGGAGGAGATCGCTTGAACCCAGGAGG	2042
Db	7195	GCACCTGTATCCAGCTACTTGGGAGGCTGAGGAGGAGATCGCTTGAACCCAGGAGG	7254
QY	2043	TGGAGGTTGCAGT	2055

Db	7255	TTGAGGTTGCAGT	7267	
RESULT 89	AC012110	141175 bp	DNA	linear
LOCUS	Homo sapiens clone RP11-45N3, WORKING DRAFT SEQUENCE, 4 unordered pieces.	141175 bp	DNA	linear
DEFINITION	AC012110	141175 bp	DNA	linear
ACCESSION	AC012110.4	GI:111336803		
VERSION	HTG; HTGS_PHASE1; HTGS_DRAFT.			
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.			
SOURCE	Homo sapiens			
ORGANISM	Homo sapiens			
REFERENCE	Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Eukaryota; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS	1 (bases 1 to 141175)			
TITLE	Birren, B., Linton, L., Nusbaum, C. and Lander, E.			
JOURNAL	Homo sapiens chromosome, clone RP11-45N3			
REFERENCE	2 (bases 1 to 141175)			
AUTHORS	Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M., Baldwin, J., Barna, N., Beckerly, R., Boguslavsky, L., Boukhgaiter, B., Brown, A., Castle, A., Collinge, M., Collins, S., Collymore, A., Cooke, P., DeArelano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M., Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D., Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Lehoczyk, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrim, J., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X., Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.			
TITLE	Direct Submission			
JOURNAL	Submitted (20-OCT-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA			
COMMENT	On Nov 10, 2000 this sequence version replaced gi:8072431. All repeats were identified using RepeatMasker: Smit, A. F. A. & Green, P. (1996-1997) http://ftp.genome.washington.edu/RM/RepeatMasker.html			
	Center: Whitehead Institute/ MIT Center for Genome Research			
	Center code: WIBR			
	Web site: http://www-seq.wi.mit.edu			
	Contact: sequence.submissions@genome.wi.mit.edu			
	----- Project Information			
	Center project name: L1203			
	Center clone name: 45_N_3			
	----- Summary Statistics			
	Sequencing vector: M13; M77815; 53% of reads			
	Sequencing vector: Plasmid; n/a; 47% of reads			
	Chemistry: Dye-primer-amersham; 2% of reads			
	Chemistry: Dye-terminator Big Dye; 98% of reads			
	Assembly program: Phrap; version 0.960731			
	Consensus quality: 140171 bases at least Q40			
	Consensus quality: 140571 bases at least Q30			
	Consensus quality: 140749 bases at least Q20			
	Insert size: 139000; agarose-1p			
	Insert size: 140875; sum-of-contigs			
	Quality coverage: 12.0 in Q20 bases.			
	* NOTE: This is a 'working draft' sequence. It currently			
	* consists of 4 contigs. The true order of the pieces			
	* is not known and their order in this sequence record is			
	* arbitrary. Gaps between the contigs are represented as			
	* runs of N, but the exact sizes of the gaps are unknown.			
	* This record will be updated with the finished sequence			
	* as soon as it is available and the accession number will			
	* be preserved.			
	* 1 93902: contig of 93902 bp in length			
	* 93903 94002: gap of 100 bp			
	* 94003 104958: contig of 10956 bp in length			
	* 104959 105058: gap of 100 bp			

FEATURES	Source	105059	122694:	contig of 17636 bp in length
		122695	122794:	gap of 100 bp
		122795	141175:	contig of 18381 bp in length.
	Location/Qualifiers			
	1..141175			
	/organism:"Homo sapiens"			
	/db_xref:"taxon:9606"			
	/clone:"Rp11-45N3"			
	/clone_lib:"RPC1-11 Human Male BAC"			
	1..93902			
	/note:"Assembly_fragment"			
	clone_end:SP6			
	vector_side:left			
	94003..104958			
	/note:"Assembly_fragment"			
	105059..122694			
	/note:"Assembly_fragment"			
	122795..141175			
	/note:"Assembly_fragment"			
	clone_end:T7			
	vector_side:right			
	BASE COUNT	42855	a	28390
	ORIGIN	c	27230	g
			42380	t
			320	others
	Query Match	2.6%	Score	73;
	Best Local Similarity	100.0%	Pred. No.	1.2e-26;
	Matches	73;	Conservative	0;
			Mismatches	0;
			Indels	0;
			Gaps	0;
QY	1983	GCACCTGTATATCCAGCTACTTTGGGAGGCTGAGGAGGAGAGTAATCGCTTGAACCCAGGAGG	2042	
Db	125489	GCACCTGTATATCCAGCTACTTTGGGAGGCTGAGGAGGAGAGTAATCGCTTGAACCCAGGAGG	125548	
QY	2043	TCGAGGTTGCAGT	2055	
Db	125549	TCGAGGTTGCAGT	125561	
RESULT 90	AC016743	144355 bp	DNA	linear
LOCUS	Homo sapiens BAC clone RP11-475H17 from 2, complete sequence.	144355 bp	DNA	linear
DEFINITION	AC016743	144355 bp	DNA	linear
ACCESSION	AC016743.10	GI:13443262		
VERSION	HTG.			
KEYWORDS	HTG.			
SOURCE	Homo sapiens.			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE	1 (bases 1 to 144355)			
AUTHORS	Sulston, J. E. and Waterston, R.			
TITLE	Toward a complete human genome sequence			
JOURNAL	Genome Res. 8 (11), 1097-1108 (1998)			
MEDLINE	99063792			
PUBMED	9847074			
REFERENCE	2 (bases 1 to 144355)			
AUTHORS	Walligorski, J., Abbott, A. and Boyer, E.			
TITLE	The sequence of Homo sapiens BAC clone RP11-475H17			
JOURNAL	Unpublished			
REFERENCE	3 (bases 1 to 144355)			
AUTHORS	Waterston, R. H.			
TITLE	Direct Submission			
JOURNAL	Submitted (04-DEC-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA			
REFERENCE	4 (bases 1 to 144355)			
AUTHORS	Waterston, R. H.			
TITLE	Direct Submission			
JOURNAL	Submitted (24-MAR-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA			
REFERENCE	5 (bases 1 to 144355)			
AUTHORS	Waterston, R.			
TITLE	Direct Submission			

## JOURNAL

Submitted (09-AUG-2001) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
6 (bases 1 to 144355)  
Waterston,R.

## REFERENCE

Direct Submission

## AUTHORS

Submitted (07-NOV-2001) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

## JOURNAL

On Mar 24, 2001 this sequence version replaced gi:13270791.  
----- Genome Center

## COMMENT

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc>

Contact: [sapiens@watson.wustl.edu](mailto:sapiens@watson.wustl.edu)

----- Summary Statistics

----- Center project name: H\_NH0475H17

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

## MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

## SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Oseogawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tatenio,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)  
VECTOR: pBACe3.6

## NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-471D6; the clone sequenced to the right is RP11-479L11, 200 bp overlap. Actual end of this clone is at base position 49393 of RP11-479L11.

## FEATURES

Location/Qualifiers

source

1..144355

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="2"

/map="2"

/clone="RP11-475H17"

/clone\_lib="RPCI-11"

405..445

/rpt\_family="AT\_rich"

1826..2062

/rpt\_family="ERVL"

3399..3600

/rpt\_family="MaLR"

4181..4244

/rpt\_family="MIR"

5716..5913

/rpt\_family="MIR"

5983..6010

/rpt\_family="AT\_rich"

6039..6272

/rpt\_family="L1"

6476..7408

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

/rpt\_family="L1"

12965..13128

/rpt\_family="L1"

13129..13445

/rpt\_family="Alu"

13446..13999

/rpt\_family="L1"

14000..14386

/rpt\_family="L1"

14387..14667

/rpt\_family="Alu"

14668..14723

/rpt\_family="L1"

14724..15471

/rpt\_family="L1"

15473..15608

/rpt\_family="L1"

15639..18141

/rpt\_family="L1"

18142..18489

/rpt\_family="MaLR"

18642..18662

/rpt\_family="AT\_rich"

20137..20494

/rpt\_family="MaLR"

20500..20566

/rpt\_family="AT\_rich"

21874..22216

/rpt\_family="MERL\_type"

22478..22536

/rpt\_family="L2"

22712..23154

/rpt\_family="ERV1"

23618..23902

/rpt\_family="L1"

24035..24055

/rpt\_family="AT\_rich"

24243..24487

/rpt\_family="MIR"

24807..25180

/rpt\_family="MaLR"

25519..25816

/rpt\_family="Alu"

25904..25966

/rpt\_family="AT\_rich"

26064..26379

/rpt\_family="Alu"

32986..33157

/rpt\_family="ERVL"

33404..33690

/rpt\_family="L1"

35170..36476

/rpt\_family="L1"

37329..37359

/rpt\_family="AT\_rich"

37537..37566

/rpt\_family="AT\_rich"

37756..38061

/rpt\_family="Alu"

38264..38503

/rpt\_family="ERV1"

38545..38585

/rpt\_family="L2"

39025..39244

/rpt\_family="ERV1"

39413..39589

/rpt\_family="ERV1"

39782..39946

/rpt\_family="MIR"

40393..40665

/rpt\_family="Alu"

41151..41855

/rpt\_family="ERV1"

```

repeat_region 41945..42232
                /rpt_family="Alu"
repeat_region 42828..43210
                /rpt_family="ERVL"
repeat_region 43211..43510
                /rpt_family="L1"
repeat_region 43511..43812
                /rpt_family="ERVL"
repeat_region 44561..44632
                /rpt_family="L2"
repeat_region 44633..45357
                /rpt_family="ERVL"
repeat_region 45358..46188
                /rpt_family="L2"
repeat_region 46279..46514
                /rpt_family="Alu"
repeat_region 46874..46958
                /rpt_family="MaLR"
repeat_region 47176..48678
                /rpt_family="L1"

```

Query Match 2.6% Score 73; DB 9; Length 144355;

```

Best Local Similarity 100.0%; Pred. No. 1.2e-26;
Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1993 TCCAGCTACTGGAGGCTGAGCGAGAGAGATCGTTGAACCCAGGAGGTGGAGTTGC 2052
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 135951 TCCAGCTACTGGAGGCTGAGCGAGAGAGATCGTTGAACCCAGGAGGTGGAGTTGC 136010
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 2053 AGTAAGCTGAGAT 2065
      |||||||||||||||
DB 136011 AGTAAGCTGAGAT 136023

```

```

RESULT 91
AC023536
LOCUS AC023536 155378 bp DNA linear HTG 09-MAY-2001
DEFINITION Homo sapiens chromosome 8 clone RP11-219D23 map 8, WORKING DRAFT
ACCESSION AC023536
VERSION AC023536.5 GI:13958545
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
1 (bases 1 to 155378)
Homo sapiens chromosome 8, clone RP11-219D23
Unpublished
2 (bases 1 to 155378)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavskiy,L.,
Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
Choepe,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Dearellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
Fenescor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
Galagan,J., Gardyna,S., Glnde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Landers,T., Largocque,K., Lehoczy,J., Levine,R.,
Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Ollivar,T.M.,
Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
Travers,M., Trigglio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
Zody,M.
Direct Submission
TITLE

```

## JOURNAL

## COMMENT

Submitted (15-FEB-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On May 6, 2001 this sequence version replaced gi:1184166.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIDR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L5446  
 Center clone name: 219.D.23  
 ----- Summary Statistics  
 Sequencing vector: M13; M7815; 2% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 153109 bases at least Q40  
 Consensus quality: 154194 bases at least Q30  
 Consensus quality: 154534 bases at least Q20  
 Insert size: 157000; agarose-fp  
 Insert size: 154678; sum-of-contigs  
 Quality coverage: 11.3 in Q20 bases; agarose-fp  
 Quality coverage: 11.4 in Q20.

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 8 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

```

1 40: contig of 40 bp in length
41 140: gap of 100 bp
141 1311: contig of 1171 bp in length
1312 1411: gap of 100 bp
1412 3166: contig of 1755 bp in length
3167 3266: gap of 100 bp
3267 59470: contig of 56204 bp in length
59471 59570: gap of 100 bp
59571 79596: contig of 20026 bp in length
79597 79696: gap of 100 bp
79697 109382: contig of 29686 bp in length
109383 109482: gap of 100 bp
109483 145165: contig of 35683 bp in length
145166 145265: gap of 100 bp
145266 155378: contig of 10113 bp in length.

```

## FEATURES

```

source
1..155378
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
/map="8"
/clone="RP11-219D23"
/clone_lib="RP11 Human Male BAC"
1..40
/feature_type="assembly_fragment"
clone_end:SP6
vector_side:left
141..1311
/feature_type="assembly_fragment"
1412..3166
/feature_type="assembly_fragment"
3267..59470
/feature_type="assembly_fragment"
59571..79596
/feature_type="assembly_fragment"
79697..109382
/feature_type="assembly_fragment"
109483..145165
/feature_type="assembly_fragment"

```



```
misc_feature      145266..155378
                    /note="assembly_fragment
                    clone_end:T7
                    vector_side:right"
BASE COUNT      46893 a 30633 c 30876 g 46276 t   700 others
ORIGIN

Query Match      2.6%   Score 73;   DB 2;   Length 155378;
Best Local Similarity 100.0%;   Pred. No. 1.12e-26;
Matches 73;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

Qy  1983 GCACCTGTAATCCAGCTACTTGGGAGGCTGAGGAGAGAAATGCTTGAACCCAGGAGG 2042
      |||
Db  116692 GCACCTGTAATCCAGCTACTTGGGAGGCTGAGGAGAGAAATGCTTGAACCCAGGAGG 116751
      |||

Qy  2043 TGGAGGTTGCAGT 2055
      |||
Db  116752 TGGAGGTTGCAGT 116764

RESULT 92
AC009305
LOCUS      AC009305      167116 bp      DNA      linear      PRI 07-OCT-2000
DEFINITION Homo sapiens BAC clone RP11-153B21 from 2, complete sequence.
ACCESSION AC009305
VERSION   AC009305.2 GI:9857572
KEYWORDS HTG.
SOURCE    Homo sapiens.
ORGANISM Homo sapiens.
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 167116)
AUTHORS   Sulston,J.E. and Waterston,R.
TITLE     Toward a complete human genome sequence
JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE   99063792
PUBMED    9847074

REFERENCE 2 (bases 1 to 167116)
AUTHORS   Kyung,K., Maupin,R., Elliot,G. and Long,N.
TITLE     The sequence of Homo sapiens BAC clone RP11-153B21
JOURNAL   Unpublished
REFERENCE 3 (bases 1 to 167116)
AUTHORS   Waterston,R.H.
TITLE     Direct Submission
JOURNAL   Submitted (13-AUG-1999) Genome Sequencing Center, Washington
           University School of Medicine, 4444 Forest Park Parkway, St. Louis,
           MO 63108, USA
REFERENCE 4 (bases 1 to 167116)
AUTHORS   Waterston,R.H.
TITLE     Direct Submission
JOURNAL   Submitted (19-AUG-2000) Genome Sequencing Center, Washington
           University School of Medicine, 4444 Forest Park Parkway, St. Louis,
           MO 63108, USA
REFERENCE 5 (bases 1 to 167116)
AUTHORS   Waterston,R.H.
TITLE     Direct Submission
JOURNAL   Submitted (10-SEP-2000) Genome Sequencing Center, Washington
           University School of Medicine, 4444 Forest Park Parkway, St. Louis,
           MO 63108, USA
REFERENCE 6 (bases 1 to 167116)
AUTHORS   Waterston,R.H.
TITLE     Direct Submission
JOURNAL   Submitted (07-OCT-2000) Department of Genetics, Washington
           University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
           On Aug 19, 2000 this sequence version replaced gi:5732166.
COMMENT    ----- Genome Center
           Center: Washington University Genome Sequencing Center
           Center code: WUGSC
           Web site: http://genome.wustl.edu/gsc
           Contact: saplens@watson.wustl.edu
           ----- Summary Statistics
           -----
           Center project name: H_NH0153B21
           -----
```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tatenoe,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACe3.6

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-541A15; the clone sequenced to the right is RP11-62F14. Actual start of this clone is at base position 1 of RP11-153B21; actual end is at base position 167116 of RP11-153B21.

FEATURES	Location/Qualifiers
source	1..167116
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="2"
	/map="2"
	/clone="RP11-153B21"
	/clone_lib="RPCI-11"
repeat_region	1483..2196
	/rpt_family="Retroviral"
repeat_region	2460..2480
	/rpt_family="(CAAAA)n"
repeat_region	3940..4032
	/rpt_family="(CA)n"
repeat_region	5461..5585
	/rpt_family="MIR"
repeat_region	5786..5815
	/rpt_family="AT-rich"
repeat_region	7025..7137
	/rpt_family="L1"
repeat_region	7178..7486
	/rpt_family="MERL_type"
repeat_region	7488..7715
	/rpt_family="L1"
repeat_region	8235..8426
	/rpt_family="MIR"
repeat_region	9134..9227
	/rpt_family="L1"
repeat_region	9924..9981
	/rpt_family="MIR"
repeat_region	10061..10155
	/rpt_family="Alu"
repeat_region	10222..10282
	/rpt_family="MIR"
repeat_region	11425..11565
	/rpt_family="MERL_type"
repeat_region	11566..11866
	/rpt_family="Alu"

```
repeat_region 11867. 11991 /rpt_family="MER1_type"
repeat_region 12108. 12263 /rpt_family="MER1_type"
repeat_region 13912. 14201 /rpt_family="Alu"
repeat_region 14588. 14891 /rpt_family="Alu"
repeat_region 14893. 15914 /rpt_family="L1"
repeat_region 16248. 16622 /rpt_family="L1"
repeat_region 16654. 16969 /rpt_family="MaLR"
repeat_region 17028. 17219 /rpt_family="L1"
repeat_region 17705. 18003 /rpt_family="Alu"
repeat_region 18232. 18533 /rpt_family="L1"
repeat_region 18740. 18966 /rpt_family="Alu"
repeat_region 19495. 19801 /rpt_family="Alu"
repeat_region 19813. 19928 /rpt_family="Other"
repeat_region 20009. 20181 /rpt_family="L1"
repeat_region 20554. 21109 /rpt_family="LTR19B"
repeat_region 23356. 23414 /rpt_family="L2"
repeat_region 25792. 25889 /rpt_family="MIR"
misc_feature 27594. 28069 /note="match to EST AA426194 (NID:92107534) zw1lf02.rl"
repeat_region 27700. 27828 /rpt_family="L1"
misc_feature 27814. 28226 /note="match to EST AI867228 (NID:g5540244) wa01f10.xl"
repeat_region 28785. 29014 /rpt_family="L1"
repeat_region 29015. 29301 /rpt_family="Alu"
repeat_region 29302. 30552 /rpt_family="L1"
repeat_region 30587. 30882 /rpt_family="Other"
repeat_region 31514. 31907 /rpt_family="L1"
repeat_region 31982. 32649 /rpt_family="L1"
repeat_region 32650. 33025 /rpt_family="MaLR"
repeat_region 33026. 33116 /rpt_family="L1"
repeat_region 33268. 33291 /rpt_family="AT-rich"
repeat_region 33382. 33635 /rpt_family="L1"
repeat_region 33683. 34038 /rpt_family="MER2_type"
repeat_region 34097. 34410 /rpt_family="L1"
repeat_region 34411. 34613 /rpt_family="MER2_type"
repeat_region 34614. 35269 /rpt_family="L1"
repeat_region 35402. 35564 /rpt_family="L1"
repeat_region 35714. 36147 /rpt_family="MER2_type"
repeat_region 37176. 37360 /rpt_family="MER2_type"

repeat_region 37939. 37974 /rpt_family="L2"
repeat_region 37939. 37974 /rpt_family="(TGG)n"

Query Match 2.6%; Score 73; DB 9; Length 167116;
Best Local Similarity 100.0%; Pred. No. 1.2e-26;
Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1983 GCACCTGTAATCCAGCTACTTGGAGGCTCAGGACGAGTAATCGTCGTGAACCCAGGAGG 2042
Db 88394 GCACCTGTAATCCAGCTACTTGGAGGCTCAGGACGAGTAATCGTCGTGAACCCAGGAGG 88453
QY 2043 TGGAGGTTGCAGT 2055
Db 88454 TGGAGGTTGCAGT 88466

RESULT 93
AC079360
LOCUS
DEFINITION Homo sapiens chromosome 12 clone RP11-7M8, WORKING DRAFT SEQUENCE,
AC079360
AC079360.22 GI:22002362
VERSION
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Aisbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinoh,K.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escottto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Homsif., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,B.E.,
Jacobson,B., Jia,Y., Johnson,R., Jollivet,S., Joudah,S.,
Karlisson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Lounsqed,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenwo,S., Oguh,M., Okwona,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ron,Y.,
Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savory,G.,
Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I.,
Sodergren,E., Sonaika,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,K., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,C., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

Direct Submission
JOURNAL
REFERENCE 2 (bases 1 to 176176)
```

AUTHORS  
TITLE  
JOURNAL

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Worley, K.C.  
Direct Submission  
Submitted (29-AUG-2000) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 1761176)  
Worley, K.C.  
Direct Submission  
Submitted (31-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jul 29, 2002 this sequence version replaced gi:19718706.

COMMENT

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Drafting Center Code: BCM  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: HCCP  
Center clone name: RP11-7M8  
----- Summary Statistics  
Sequencing vector: Plasmid;  
Sequencing vector: M13;  
Chemistry: Dye-Primer Bodipy: 18% of reads  
Chemistry: Dye-Terminator Big Dye: 82% of reads  
Assembly program: Phrap: version 0.990329  
Consensus quality: 169109 bases at least Q40  
Consensus quality: 171607 bases at least Q30  
Consensus quality: 173502 bases at least Q20  
Estimated insert size: 163861; sum-of-contigs estimation  
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation  
-----

\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 12 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1  
2163: contig of 2163 bp in length  
2263: gap of unknown length  
2264: 4836: contig of 2573 bp in length  
4837: 4936: gap of unknown length  
4937: 7285: contig of 2349 bp in length  
7286: 7385: gap of unknown length  
7386: 9800: contig of 2415 bp in length  
9801: 14017: contig of 4117 bp in length  
14018: 14117: gap of unknown length  
14118: 19883: contig of 5766 bp in length  
19884: 19984: gap of unknown length  
19984: 25706: contig of 5723 bp in length  
25707: 25806: gap of unknown length  
25807: 33908: contig of 8102 bp in length  
33909: 34008: gap of unknown length  
34009: 42136: contig of 8128 bp in length  
42137: 42237: gap of unknown length  
42237: 59945: contig of 17709 bp in length  
59946: 60045: gap of unknown length  
60046: 82634: contig of 22589 bp in length  
82635: 82734: gap of unknown length  
82735: 176176: contig of 93442 bp in length.

FEATURES  
source

1..176176  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="12"  
/clone="RP11-7M8"

BASE COUNT

40817 a 46113 c 45951 g 42181 t 1114 others

ORIGIN

Query Match 2.6%; Score 73; DB 2; Length 176176;  
Best Local Similarity 100.0%; Pred. No. 1.2e-26;  
Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1983 GCACCTGTAATCCCGAGCTACTTGGGAGGCTGAGGAGGAGGAGTGCCTTGAACCCAGGAGG 2042  
|||||  
Db 60822 GCACCTGTAATCCCGAGCTACTTGGGAGGCTGAGGAGGAGGAGTGCCTTGAACCCAGGAGG 60881  
|||||  
  
QY 2043 TGGAGGTTGCAGT 2055  
|||||  
Db 60882 TGGAGGTTGCAGT 60894  
|||||

RESULT 94  
AL162211/c

LOCUS  
DEFINITION Homo sapiens chromosome 13 clone RP11-157H17, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*, 16 unordered pieces..

ACCESSION AL162211  
VERSION AL162211.5 GI:9863623  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_CANCELLED.  
SOURCE human.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 179470)  
Burton, J.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)  
requests: [clonerequest@sanger.ac.uk](mailto:clonerequest@sanger.ac.uk)  
On Aug 21, 2000 this sequence version replaced gi:9212905.  
----- Genome Center  
Center: Sanger Centre  
Center code: SC  
Web site: <http://www.sanger.ac.uk>  
Contact: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)  
----- Project Information  
Center project name: BA157H17  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Sequencing vector: plasmid; 108752; 100% of reads  
Chemistry: Dye-terminator Big Dye: 100% of reads  
Consensus quality: 170263 bases at least Q40  
Consensus quality: 173209 bases at least Q30  
Consensus quality: 175040 bases at least Q20  
Insert size: 177970; sum-of-contigs  
Insert size: 218642; 8.2% error; agarose-fp  
Quality coverage: 3.74x in Q20 bases; sum-of-contigs Quality  
coverage: 3.12x in Q20 bases; agarose-fp  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 16 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 20111: contig of 20111 bp in length  
\* 20112 20211: gap of 100 bp  
\* 20212 22684: contig of 2473 bp in length  
\* 22685 22784: gap of 100 bp  
\* 22785 26050: contig of 3266 bp in length  
\* 26051 26150: gap of 100 bp  
\* 26151 50870: contig of 24720 bp in length  
\* 50871 50970: gap of 100 bp  
\* 50971 56714: contig of 5744 bp in length  
\* 56715 56814: gap of 100 bp  
\* 56815 79207: contig of 22392 bp in length  
\* 79207 79306: gap of 100 bp

\* 79307 82214: contig of 2908 bp in length  
 \* 82215 82314: gap of 100 bp  
 \* 82315 86916: contig of 4602 bp in length  
 \* 86917 87016: gap of 100 bp  
 \* 87017 98674: contig of 11658 bp in length  
 \* 98675 98774: gap of 100 bp  
 \* 98775 112648: contig of 13874 bp in length  
 \* 112649 112748: gap of 100 bp  
 \* 112749 126894: contig of 14146 bp in length  
 \* 126895 126994: gap of 100 bp  
 \* 126995 140436: contig of 13442 bp in length  
 \* 140437 140536: gap of 100 bp  
 \* 140537 148268: contig of 7732 bp in length  
 \* 148269 148368: gap of 100 bp  
 \* 148369 151247: contig of 2879 bp in length  
 \* 151248 151347: gap of 100 bp  
 \* 151348 160829: contig of 9482 bp in length  
 \* 160830 160929: gap of 100 bp  
 \* 160930 179470: contig of 18541 bp in length.

## FEATURES

source

Location/Qualifiers  
 1. .179470  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="13"  
 /clone="RP11-157H17"  
 /clone\_lib="RPC1-11.1"  
 1. .20111  
 /note="assembly\_fragment:01022  
 fragment\_chain:1"  
 20212..22684  
 /note="assembly\_fragment:00962  
 fragment\_chain:1"  
 22785..26050  
 /note="assembly\_fragment:00754  
 fragment\_chain:1"  
 26151..50870  
 /note="assembly\_fragment:00208  
 fragment\_chain:2"  
 50971..56714  
 /note="assembly\_fragment:01315  
 fragment\_chain:2"  
 56815..79206  
 /note="assembly\_fragment:00254  
 fragment\_chain:3"  
 79307..82214  
 /note="assembly\_fragment:00835  
 fragment\_chain:3"  
 82315..86916  
 /note="assembly\_fragment:00293  
 fragment\_chain:4"  
 87017..98674  
 /note="assembly\_fragment:00855  
 fragment\_chain:4"  
 98775..112648  
 /note="assembly\_fragment:01017  
 fragment\_chain:5"  
 112749..126894  
 /note="assembly\_fragment:01010  
 fragment\_chain:5"  
 126895..140436  
 /note="assembly\_fragment:01132  
 fragment\_chain:6"  
 140537..148268  
 /note="assembly\_fragment:00107  
 fragment\_chain:6"  
 148369..151247  
 /note="assembly\_fragment:00075"  
 151348..160829  
 /note="assembly\_fragment:00269"  
 160930..179470  
 /note="assembly\_fragment:00989"  
 179470..18541

BASE COUNT 54009 a 35515 c 34477 g 53955 t 1514 others  
 ORIGIN

Query Match 2.6%: Score 73; DB 2: Length 179470;  
 Best Local Similarity 100.08; Pred. No. 1.2e-26;  
 Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1983 GCACCTGTAATCCCGAGCTTGGAGGCTGAGCAGGAGAGATCGCTTGAAGCCAGGAGG 2042  
 Db 95945 GCACCTGTAATCCCGAGCTTGGAGGCTGAGCAGGAGAGATCGCTTGAAGCCAGGAGG 95886  
 QY 2043 TGGAGGTTGCCAGT 2055  
 Db 95885 TGGAGGTTGCCAGT 95873  
 RESULT 95  
 AC021915/c  
 LOCUS 183101 bp DNA linear PRI 40-APR-2002  
 DEFINITION Homo sapiens chromosome 8, clone RP11-26M5, complete sequence.  
 ACCESSION AC021915  
 VERSION AC021915.11 GI:20128647  
 KEYWORDS HTG  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 183101)  
 Homo sapiens chromosome 8, clone RP11-26M5  
 Unpublished  
 2 (bases 1 to 183101)  
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
 Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Bada, F.,  
 Boguslavskiy, L., Boukhgalter, B., Brown, A., Burkett, G., Castle, A.,  
 Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,  
 Dearellano, K., Dewar, K., Domino, M., Doyle, M., Fenebor, J.,  
 Ferreira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J.,  
 Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,  
 Howland, J. C., Johnson, R., Jones, C., Kann, L., Karas, A., Klein, J.,  
 Landers, T., Lechovsky, J., Levine, R., Liu, C., Liu, G., Locke, K.,  
 Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K.,  
 McPheters, R., Meldrim, J., Menus, L., Morrow, J., Naylor, J.,  
 Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K.,  
 Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,  
 Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,  
 Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
 Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,  
 Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (22-JAN-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 183101)  
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,  
 Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L.,  
 Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,  
 Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,  
 Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S.,  
 Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,  
 Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
 Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
 Kamat, A., Karas, A., Kells, C., LaRocque, K., Lamazatus, R.,  
 Landers, T., Lechovsky, J., Levine, R., Lindblad-Toh, K., Liu, G.,  
 Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,  
 McCarthy, M., McEwan, P., McKernan, K., Meldrim, J., Menus, L.,  
 Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,  
 Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,  
 Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,  
 Raymond, C., Retta, R., Riback, M., Riley, R., Riese, C., Rios, P.,  
 Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, K.,  
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
 Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
 Topham, K., Travers, M., Travis, N., Trigglio, J., Vassiliev, H.,  
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,  
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.





repeat\_region /note="MADE1 repeat: matches 1. .80 of consensus"  
44496. .44931  
repeat\_region /note="L2 repeat: matches 2112. .2572 of consensus"  
45513. .47217  
repeat\_region /note="L1PA6 repeat: matches 4434. .6143 of consensus"  
47218. .47396  
repeat\_region /note="L1PB3 repeat: matches 5956. .6126 of consensus"  
50159. .51245  
repeat\_region /note="L2 repeat: matches 1609. .2750 of consensus"  
51426. .51726  
repeat\_region /note="L2 repeat: matches 1199. .1530 of consensus"  
51733. .51845  
repeat\_region /note="L1TR29 repeat: matches 2. .118 of consensus"  
51773. .52349  
repeat\_region /note="MER39b repeat: matches 4. .577 of consensus"  
52355. .52386  
repeat\_region /note="16 copies 2 mer aa 90% conserved"  
52499. .52801  
repeat\_region /note="AluSp repeat: matches 1. .313 of consensus"  
53175. .53430  
repeat\_region /note="L2 repeat: matches 1. .256 of consensus"  
53486. .53621  
repeat\_region /note="L2 repeat: matches 1617. .1753 of consensus"  
55490. .56705  
repeat\_region /note="L1MA5 repeat: matches 5064. .6300 of consensus"  
56706. .57003  
repeat\_region /note="AluSq repeat: matches 1. .304 of consensus"  
57004. .57087  
repeat\_region /note="L1MA5 repeat: matches 4980. .5064 of consensus"  
57088. .57374  
repeat\_region /note="AluJo repeat: matches 1. .284 of consensus"  
57375. .58079  
repeat\_region /note="L1MA5 repeat: matches 4339. .4980 of consensus"  
58080. .58374  
repeat\_region /note="AluSk repeat: matches 1. .294 of consensus"  
58375. .58458  
repeat\_region /note="L1MA5 repeat: matches 4255. .4339 of consensus"  
58489. .59221  
repeat\_region /note="L1PA12 repeat: matches 5411. .6153 of consensus"  
59217. .60773  
repeat\_region /note="L1MA5 repeat: matches 2675. .4268 of consensus"  
60988. .61059  
repeat\_region /note="L1M2 repeat: matches 354. .423 of consensus"  
61512. .62087  
repeat\_region /note="L1M2 repeat: matches -668. -.100 of consensus"  
62833. .63254  
repeat\_region /note="NSTC repeat: matches 3. .405 of consensus"  
64230. .64526  
repeat\_region /note="AluSg repeat: matches 1. .296 of consensus"  
66309. .66499  
repeat\_region /note="L1M02 repeat: matches 6154. .6340 of consensus"  
67079. .67376  
repeat\_region /note="AluSk repeat: matches 2. .304 of consensus"  
67502. .67541  
repeat\_region /note="10 copies 4 mer atat 82% conserved"  
67822. .67838  
repeat\_region /note="L2 repeat: matches 2732. .2749 of consensus"  
67839. .68126  
repeat\_region /note="AluJo repeat: matches 14. .300 of consensus"  
68127. .68257  
repeat\_region /note="L2 repeat: matches 2616. .2732 of consensus"  
70069. .70100  
repeat\_region /note="16 copies 2 mer gt 84% conserved"  
70410. .73012  
repeat\_region /note="L1M1 repeat: matches 3067. .5737 of consensus"  
73878. .73905  
repeat\_region /note="14 copies 2 mer aa 89% conserved"

Query Match 2.6%; Score 73; DB 9; Length 191652;  
Best Local Similarity 100.0%; Pred. No. 1.2e-26;  
Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 188583 GCACCTGTAATCCAGCTACTTGGGAGGCTGAGGAGGAGTAATCGCTTGAACCCAGAGG 188524  
Oy 2043 TGGAGGTTGCAGT 2055  
Db 188523 TGGAGGTTGCAGT 188511  
RESULT 97  
AC017100 202950 bp DNA linear PRI 25-JAN-2002  
LOCUS Homo sapiens BAC clone RP11-549B18 from 18, complete sequence.  
DEFINITION AC017100  
ACCESSION AC017100  
VERSION AC017100.4 GI:11120958  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 202950)  
AUTHORS Sulston,J.E. and Waterston,R.  
TITLE Toward a complete human genome sequence  
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)  
MEDLINE 99063792  
PUBMED 9847074  
REFERENCE 2 (bases 1 to 202950)  
AUTHORS Edwards,J., Gregory,S., Maupin,R. and O'Rourke,K.  
TITLE The sequence of Homo sapiens BAC clone RP11-549B18  
JOURNAL Unpublished (2002)  
REFERENCE 3 (bases 1 to 202950)  
AUTHORS Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (09-DEC-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
REFERENCE 4 (bases 1 to 202950)  
AUTHORS Waterston,R.  
TITLE Direct Submission  
JOURNAL Submitted (08-NOV-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
REFERENCE 5 (bases 1 to 202950)  
AUTHORS Waterston,R.  
TITLE Direct Submission  
JOURNAL Submitted (09-MAY-2001) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
REFERENCE 6 (bases 1 to 202950)  
AUTHORS Waterston,R.  
TITLE Direct Submission  
JOURNAL Submitted (25-JAN-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
COMMENT On Nov 8, 2000 this sequence version replaced gi:7631062.  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: http://genome.wustl.edu/gsc  
Contact: saplens@watson.wustl.edu  
----- Summary Statistics  
-----  
Center project name: H\_NH0549B18

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.





Qy	2043	TGGAGGTTGCAGT	2055
Db	36169	TGGAGGTTGCAGT	36181
RESULT 98			
AP001336/c			
LOCUS			
DEFINITION	AP001336 Homo sapiens chromosome 18 clone RP11-735F7 map 18q12, WORKING DRAFT SEQUENCE, 18 unordered pieces.		
ACCESSION	AP001336.3 GI:9857606		
VERSION	HTG: HTGS_PHRASE1; HTGS_DRAFT..		
KEYWORDS	Homo sapiens DNA, clone:RP11-735F7.		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 (bases 1 to 216441)		
AUTHORS	Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.		
TITLE	Homo sapiens 216,441 genomic DNA of 18q12		
JOURNAL	Published Only in DataBase (2000)		
REFERENCE	2 (bases 1 to 216441)		
AUTHORS	Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.		
TITLE	Direct Submission		
JOURNAL	Submitted (02-MAR-2000) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-15-1 Kitasato, Sagami-hara, Kanagawa 228-8555, Japan		
COMMENT	(E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923, Fax:81-42-778-9924) On Aug 19, 2000 this sequence version replaced gi:8117259. ----- Genome Center Center: RIKEN Genomic Sciences Center(GSC) Center code: RIKEN Web site: http://hgp.gsc.riken.go.jp/ Contact: hattori@gsc.riken.go.jp ----- Project Information Center project name: Humdraft18 Center clone name: RP11-735F7 ----- Summary Statistics Sequencing vector: PCR products; 100% of reads Chemistry: Dye-terminator ET-amersham; 100% of reads Assembly program: Phrap; version 0.990329 Consensus quality: 210841 bases at least Q40 Consensus quality: 212964 bases at least Q30 Consensus quality: 214061 bases at least Q20 Insert size: 214741; sum-of-contigs Quality coverage: 10.84x in Q20 bases; sum-of-contigs ----- NOTE: This is a 'working draft' sequence. It currently consists of 18 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.  1 35982 contig of 35982 bp in length 36083 65042 contig of 28960 bp in length 65143 92388 contig of 2746 bp in length 92489 116561 contig of 24073 bp in length 116662 137695 contig of 21034 bp in length 137796 154184 contig of 16389 bp in length 154285 166582 contig of 12298 bp in length 166683 178123 contig of 11441 bp in length 178224 184709 contig of 6486 bp in length 184810 191831 contig of 7022 bp in length 191932 198683 contig of 6752 bp in length 198784 202513 contig of 3730 bp in length 202614 205890 contig of 3277 bp in length 205991 209826 contig of 3836 bp in length 209927 212133 contig of 2207 bp in length 212234 213778 contig of 1545 bp in length 213879 215264 contig of 1386 bp in length 215265 215364 gap of 100 bp 215365 216441 contig of 1077 bp in length.  FEATURES Source 1..216441 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="18" /map="18q12" /clone="RP11-735F7" 1..35982 /note="assembly_fragment clone_end:SP6_vector_side:left" 36083..65042 /note="assembly_fragment" 65143..92388 /note="assembly_fragment" 92489..116561 /note="assembly_fragment" 116662..137695 /note="assembly_fragment" 137796..154184 /note="assembly_fragment" 154285..166582 /note="assembly_fragment" 166683..178123 /note="assembly_fragment" 178224..184709 /note="assembly_fragment clone_end:T7_vector_side:left" 184810..191831 /note="assembly_fragment" 191932..198683 /note="assembly_fragment"		

```

/note="assembly_fragment"
198784..202513
/note="assembly_fragment"
202614..205890
/note="assembly_fragment"
205991..209826
/note="assembly_fragment"
209927..212133
/note="assembly_fragment"
212234..213778
/note="assembly_fragment"
213879..215264
/note="assembly_fragment"
215365..216441
/note="assembly_fragment"
216442..217000

BASE COUNT 63565 a 42207 c 43579 g 65390 t 1700 others
ORIGIN

Query Match      2.6%; Score 73; DB 2; Length 216441;
Best Local Similarity 100.0%; Pred. No. 1.1e-26;
Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1983 GCACCTGTATCCAGCTACTTGGAGGCTTGAGCAGGAGATCGCTTGAACCCAGGAG 2042
      |||||
Db 124755 GCACCTGTATCCAGCTACTTGGAGGCTTGAGCAGGAGATCGCTTGAACCCAGGAG 124696
      |||||

QY 2043 TGGAGGTTCAGT 2055
      |||||
Db 124695 TGGAGGTTCAGT 124683

RESULT 99
AC022460/c
LOCUS
DEFINITION      Homo sapiens chromosome 12 clone RP11-403N24, WORKING DRAFT
ACCESSION      AC022460
VERSION        AC022460.4 GI:8954357
KEYWORDS       HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE         Homo sapiens.
ORGANISM       Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 239008)
Waterston,R.H.
Unpublished
2 (bases 1 to 239008)
Waterston,R.H.
Direct Submission
Submitted (03-FEB-2000) Genome Sequencing Center, Washington
University School of Medicine, 444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Jul 7, 2000 this sequence version replaced gi:8570009.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NR0403N24
----- Summary Statistics -----
Sequencing vector: M13; 75%
Chemistry: Dye-primer ET; 75% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 223186 bases at least Q40
Consensus quality: 228280 bases at least Q30
Consensus quality: 230978 bases at least Q20
Insert size: 212000; agarose-fp
Insert size: 236308; sum-of-contigs
Quality coverage: 5.72 in Q20 bases; agarose-fp

```

```

Quality coverage: 5.64 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 28 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 1012: contig of 1012 bp in length
* 1013: gap of unknown length
* 1113: contig of 1227 bp in length
* 2339: gap of unknown length
* 2439: gap of unknown length
* 3628: contig of 1189 bp in length
* 3728: gap of unknown length
* 4968: contig of 1240 bp in length
* 5068: gap of unknown length
* 6837: contig of 1769 bp in length
* 6937: gap of unknown length
* 8331: contig of 1394 bp in length
* 8332: gap of unknown length
* 8432: contig of 1030 bp in length
* 9561: gap of unknown length
* 9562: contig of 3264 bp in length
* 12825: gap of unknown length
* 12826: gap of unknown length
* 14654: gap of unknown length
* 14754: contig of 1998 bp in length
* 16752: gap of unknown length
* 16851: contig of 2504 bp in length
* 19355: gap of unknown length
* 19455: gap of unknown length
* 22127: contig of 2672 bp in length
* 22227: gap of unknown length
* 26873: contig of 4646 bp in length
* 26973: gap of unknown length
* 32654: contig of 5681 bp in length
* 32754: gap of unknown length
* 38338: contig of 5584 bp in length
* 38438: gap of unknown length
* 47366: contig of 8928 bp in length
* 47367: gap of unknown length
* 47467: contig of 8054 bp in length
* 55620: gap of unknown length
* 55621: contig of 6252 bp in length
* 61872: gap of unknown length
* 61973: gap of unknown length
* 68095: contig of 6123 bp in length
* 68195: gap of unknown length
* 75555: contig of 7360 bp in length
* 75556: gap of unknown length
* 83079: contig of 7424 bp in length
* 83080: gap of unknown length
* 83180: contig of 11865 bp in length
* 95044: gap of unknown length
* 95144: gap of unknown length
* 95145: contig of 12103 bp in length
* 107248: gap of unknown length
* 107348: contig of 14214 bp in length
* 121562: gap of unknown length
* 121662: contig of 16468 bp in length
* 138129: gap of unknown length
* 138229: gap of unknown length
* 156911: contig of 18682 bp in length
* 156912: gap of unknown length
* 157012: contig of 27148 bp in length
* 184160: gap of unknown length
* 184260: contig of 54749 bp in length.
*
* Location/Qualifiers
* 1. 239008
* /organism="Homo sapiens"
* /db_xref="taxon:9606"
* /chromosome="12"
* /clone="RP11-403N24"
*
* BASE COUNT 59374 a 57881 c 58172 g 60872 t 2709 others
* ORIGIN

```

Query Match 2.6%; Score 73; DB 2; Length 239008;  
Best Local Similarity 100.0%; Pred. No. 1.2e-26;  
Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1983 GCACCTGTAATCCAGCTACTTGGAGGCTGAGGAGGAGTAATCGCTTGAACCCAGGAGG 2042  
|||||  
DB 61157 GCACCTGTAATCCAGCTACTTGGAGGCTGAGGAGGAGTAATCGCTTGAACCCAGGAGG 61098  
|||||

QY 2043 TGGAGGTTGCAGT 2055  
|||||

DB 61097 TGGAGGTTGCAGT 61085  
|||||

Db 42497 CACCTGTAATCCAGCTACTTGGAGGCTGAGGAGGAGTAATCGCTTGAACCCAGGAGGT 42556  
QY 2044 GGAGGTTGCAGT 2055  
|||||  
Db 42557 GGAGGTTGCAGT 42568  
|||||

Search completed: July 18, 2003, 01:11:12  
Job time : 7146 secs

RESULT 100  
AL669821 89211 bp DNA linear PRI 24-APR-2002  
LOCUS Human DNA sequence from clone XHbac-116A1 on chromosome 6, complete  
DEFINITION sequence.  
ACCESSION AL669821  
VERSION AL669821.9 GI:20152403  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1  
AUTHORS Mashreghi-Mohammadi, M.  
TITLE Direct Submission  
JOURNAL Submitted (23-APR-2002) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquerry@sanger.ac.uk  
On Apr 15, 2002 this sequence version replaced gi:20145365.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >= 30);  
an attempt was made to resolve all sequencing problems, such as  
compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. The following  
abbreviations are used to associate primary accession numbers given  
in the feature table with their source databases: Em., EMBL; Sw.,  
SWISSPROT; Tr., TrEMBL; Wp., WormPep; Information on the WormPep  
database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep  
from a CHORI-501 human bac - PGF cell line library VECTOR:  
PTARBAC2.1  
This sequence was generated from part of bacterial clone contigs  
constructed by the MHC Haplotype Consortium and collaborators.  
Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr6/MHC.  
FEATURES  
Location/Qualifiers  
1..89211  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/clone="XHbac-116A1"  
/clone\_lib="CHORI-501"  
BASE COUNT 26990 a 19373 c 19477 g 23371 t  
ORIGIN  
Query Match 2.6%; Score 72; DB 9; Length 89211;  
Best Local Similarity 100.0%; Pred. No. 4.1e-26;  
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1984 CACCTGTAATCCAGCTACTTGGAGGCTGAGGAGGAGTAATCGCTTGAACCCAGGAGGT 2043  
|||||



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 17, 2003, 20:58:25 ; Search time 1772 seconds  
(without alignments)  
3251.894 Million cell updates/sec

Title: US-09-966-880A-8

Perfect score: 198

Sequence: 1 MDSLLMNRKFLYQPKNVRW.....ILLPLYEVDLDRFAFTLGL 198

Scoring table:  
OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Word size: 1

Total number of hits satisfying chosen parameters: 4106490

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 500 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlp  
-O=/cn2.1/USPTO.scoo1/US09966880/runat\_16072003\_115617.17474/app\_query.fasta\_1.391  
-DB=GenEmbl -QFWT=fastap -SUFFIX=10align.rge -MINMATCH=0.1 -LOOPCT=0  
-LOOPEXT=0 -DUALS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=500 -DOALIGN=200 -THR\_SCORE=quality -THR\_MIN=1 -ALIGN=100 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09966880.ecgn\_1.1.3745@runat\_16072003\_115617.17474 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : GenEmbl :  
1: gb\_ba.\*  
2: gb\_htg.\*  
3: gb\_in.\*  
4: gb\_ov.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_on.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*

29: em\_vi.\*  
30: em\_htg\_hum.\*  
31: em\_htg\_inv.\*  
32: em\_htg\_other.\*  
33: em\_htg\_mus.\*  
34: em\_htg\_pln.\*  
35: em\_htg\_rtd.\*  
36: em\_htg\_mam.\*  
37: em\_htg\_vrt.\*  
38: em\_sy.\*  
39: em\_hgo\_hum.\*  
40: em\_hgo\_mus.\*  
41: em\_hgo\_other.\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	198	100.0	596	9 AF529823	AF529823 Homo sapi
2	198	100.0	596	9 AF529826	AF529826 Homo sapi
3	198	100.0	596	12 AF529837	AF529837 Mus muscu
4	198	100.0	1837	9 BC006296	BC006296 Homo sapi
5	198	100.0	2791	9 AB040431	AB040431 Homo sapi
6	195	98.5	596	12 AF529829	AF529829 Mus muscu
7	195	98.5	597	12 AF529842	AF529842 Crictulu
8	195	98.5	597	12 AF529843	AF529843 Crictulu
9	195	98.5	597	12 AF529844	AF529844 Crictulu
10	195	98.5	597	12 AF529847	AF529847 Crictulu
11	195	98.5	597	12 AF529852	AF529852 Crictulu
12	190	96.0	597	12 AF529846	AF529846 Crictulu
13	177	89.4	596	12 AF529832	AF529832 Mus muscu
14	175	88.4	596	9 AF529815	AF529815 Homo sapi
15	175	88.4	596	12 AF529831	AF529831 Mus muscu
16	172	86.9	597	12 AF529851	AF529851 Crictulu
17	171	86.4	596	9 AF529825	AF529825 Homo sapi
18	170	85.9	591	12 AF529855	AF529855 Crictulu
19	169	85.4	596	12 AF529833	AF529833 Mus muscu
20	169	85.4	596	12 AF529835	AF529835 Mus muscu
21	162	81.8	547	12 AF529856	AF529856 Crictulu
22	160	80.8	596	9 AF529822	AF529822 Homo sapi
23	149	75.3	595	12 AF529834	AF529834 Mus muscu
24	146	73.7	596	12 AF529836	AF529836 Mus muscu
25	146	73.7	597	12 AF529853	AF529853 Crictulu
26	144	72.7	597	12 AF529828	AF529828 Mus muscu
27	143	72.2	596	9 AF529824	AF529824 Homo sapi
28	140	70.7	597	12 AF529841	AF529841 Crictulu
29	139	70.2	596	12 AF529830	AF529830 Mus muscu
30	136	68.7	596	12 AF529838	AF529838 Mus muscu
31	135	68.2	597	12 AF529845	AF529845 Crictulu
32	128	64.6	596	9 AF529821	AF529821 Homo sapi
33	126	63.6	597	12 AF529848	AF529848 Crictulu
34	122	61.6	596	12 AF529839	AF529839 Mus muscu
35	122	61.6	597	12 AF529850	AF529850 Crictulu
36	118	59.6	596	9 AF529818	AF529818 Homo sapi
37	118	59.6	596	9 AF529819	AF529819 Homo sapi
38	116	58.6	596	9 AF529820	AF529820 Homo sapi
39	115	58.6	597	12 AF529849	AF529849 Crictulu
40	115	58.1	577	12 AF529840	AF529840 Mus muscu
41	114	57.6	596	9 AF529816	AF529816 Homo sapi
42	109	55.1	596	9 AF529827	AF529827 Homo sapi
43	108	54.5	597	12 AF529854	AF529854 Crictulu
44	96	48.5	596	9 AF529817	AF529817 Homo sapi
45	90	45.5	11204	9 AB040430	AB040430 Homo sapi
46	90	45.5	71132	9 AC092184	AC092184 Homo sapi
47	42	21.2	2440	10 AF132979	AF132979 Mus muscu
48	42	21.2	143044	2 AC094826	AC094826 Rattus no
49	42	21.2	177103	2 AC120617	AC120617 Rattus no
50	38	19.2	50269	2 AC119975	AC119975 Mus muscu

C 51	34	17.2	98161	2	AC109119	AC109119 Rattus no	124	8	4.0	2086	8	AF224075	AF224075 Solanum n
C 52	16	8.1	50269	2	AC119975	Mus muscu	125	8	4.0	2086	8	CBU08916	CBU08916 Capsicum ba
C 53	10	5.1	67244	2	AC068309	Mus muscu	126	8	4.0	2086	8	CBU47428	CBU47428 Cyphomandra
C 54	9	4.5	7750	3	DMSPALTR	Y07653 D.melanogas	127	8	4.0	2086	8	DSU08917	DSU08917 Datura stra
C 55	55	4.5	10945	1	AE007918	AE007918 Agrobacte	128	8	4.0	2086	8	JAU08919	JAU08919 Juanelloa a
C 56	9	4.5	11013	1	AE008970	AE008970 Agrobacte	129	8	4.0	2086	8	JPU47429	JPU47429 Jalomata p
C 57	9	4.5	59915	2	AC100984	Mus muscu	130	8	4.0	2086	8	LCU08920	LCU08920 Lycium cest
C 58	9	4.5	103746	2	AC014357	AC014357 Drosophil	131	8	4.0	2086	8	LEU08921	LEU08921 Lycium cest
C 59	9	4.5	141899	2	AC087457	AC087457 Homo sapi	132	8	4.0	2086	8	LEU73797	LEU73797 Lycianthes
C 60	9	4.5	150147	2	AC123206	AC123206 Rattus no	133	8	4.0	2086	8	MOU08922	MOU08922 Mandragora
C 61	9	4.5	154288	2	AL513164	AL513164 Human DNA	134	8	4.0	2086	8	NAU08923	NAU08923 Nicotiana a
C 62	9	4.5	158016	2	AC113662	AC113662 Rattus no	135	8	4.0	2086	8	NP008924	NP008924 Nicotiana a
C 63	9	4.5	173509	2	AP002436	AP002436 Homo sapi	136	8	4.0	2086	8	NSU08925	NSU08925 Nicotiana ph
C 64	9	4.5	176053	2	AC068190	AC068190 Homo sapi	137	8	4.0	2086	8	PAU08927	PAU08927 Physalis al
C 65	9	4.5	179510	2	AC127421	AC127421 Mus muscu	138	8	4.0	2086	8	POTCP1A	POTCP1A Physalis al
C 66	9	4.5	180668	2	AC020857	AC020857 Mus muscu	139	8	4.0	2086	8	SAU47417	SAU47417 Solanum tor
C 67	9	4.5	187948	2	AP003402	AP003402 Homo sapi	140	8	4.0	2086	8	SAU47418	SAU47418 Solanum tub
C 68	9	4.5	188067	3	AC007083	AC007083 Drosophil	141	8	4.0	2086	8	SAU72752	SAU72752 Solanum arb
C 69	9	4.5	188136	2	AL840638	AL840638 Dario rer	142	8	4.0	2086	8	SGU72751	SGU72751 Solanum avi
C 70	9	4.5	190960	2	AC036188	AC036188 Homo sapi	143	8	4.0	2086	8	SGU72752	SGU72752 Solanum arg
C 71	9	4.5	199386	2	NCB14D6	NCB14D6 Neurospor	144	8	4.0	2086	8	SGU72753	SGU72753 Solanum cor
C 72	9	4.5	207584	2	AC103892	AC103892 Rattus no	145	8	4.0	2086	8	SPU47421	SPU47421 Solanum gr
C 73	9	4.5	232951	2	AC091322	AC091322 Mus muscu	146	8	4.0	2086	8	SPU47422	SPU47422 Solanum glia
C 74	9	4.5	24237	3	AE003632	AE003632 Drosophil	147	8	4.0	2086	8	SPU47423	SPU47423 Solanum phy
C 75	8	4.0	245	8	TCART1S1	TCART1S1 Trautvetter	148	8	4.0	2086	8	SRU47424	SRU47424 Solanum pse
C 76	8	4.0	354	9	HS2A279528	HS2A279528 Homo sapi	149	8	4.0	2086	8	SSU47425	SSU47425 Solanum ros
C 77	8	4.0	500	6	AX380702	AX380702 Sequence	150	8	4.0	2086	8	STU72750	STU72750 Solanum sea
C 78	8	4.0	569	11	HSC07H12	AL158650 H.sapiens	151	8	4.0	2086	8	STU72751	STU72751 Solanum t
C 79	8	4.0	653	11	PM3C3B	AL685634 Penicilli	152	8	4.0	2086	8	SWU47426	SWU47426 Solanum wal
C 80	8	4.0	984	6	A48781	A48781 Sequence 8	153	8	4.0	2086	8	WSU72755	WSU72755 Solanum t
C 81	8	4.0	984	6	A58739	A58739 Sequence 2	154	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 82	8	4.0	984	6	AR018707	AR018707 Sequence	155	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 83	8	4.0	1006	6	A48778	A48778 Sequence 5	156	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 84	8	4.0	1006	6	AR018705	AR018705 Sequence	157	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 85	8	4.0	1101	3	AY094934	AY094934 Drosophil	158	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 86	8	4.0	1149	6	A48779	A48779 Sequence 6	159	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 87	8	4.0	1149	6	A48782	A48782 Sequence 9	160	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 88	8	4.0	1149	6	AR018706	AR018706 Sequence	161	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 89	8	4.0	1149	6	AR018708	AR018708 Sequence	162	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 90	8	4.0	1188	6	A58738	A58738 Sequence 1	163	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 91	8	4.0	1230	6	A58740	A58740 Sequence 3	164	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 92	8	4.0	1237	8	AF238055	AF238055 Hydrastis	165	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 93	8	4.0	1237	8	AF060158	AF060158 Aucuba ja	166	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 94	8	4.0	1465	1	AF071567	AF071567 Bordetell	167	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 95	8	4.0	1565	8	EANDHF1	AF147710 Exacum af	168	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 96	8	4.0	1857	1	AF353205	AF353205 Azospiril	169	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 97	8	4.0	1886	1	AK094292	AK094292 Homo sapi	170	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 98	8	4.0	1938	8	HAU12660	HAU12660 Hemigraphis	171	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 99	8	4.0	1938	8	AF257485	AF257485 Goyazia r	172	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 100	8	4.0	2025	8	ESU12659	U12659 Eremomastax	173	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 101	8	4.0	2052	8	AF206197	AF206197 Nematanth	174	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 102	8	4.0	2057	8	AF013689	AF013689 Nematanth	175	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 103	8	4.0	2057	8	AF206198	AF206198 Nematanth	176	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 104	8	4.0	2057	8	AF206196	AF206196 Codonanth	177	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 105	8	4.0	2063	8	AF206196	AF206196 Tylopsaca	178	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 106	8	4.0	2067	8	AF176635	AF176635 Solanum m	179	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 107	8	4.0	2071	8	AF224068	AF224068 Solanum m	180	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 108	8	4.0	2077	8	AF040144	AF040144 Bellonia	181	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 109	8	4.0	2080	8	SAU47415	U47415 Solanum abu	182	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 110	8	4.0	2084	8	SLU47420	U47420 Solanum lac	183	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 111	8	4.0	2086	8	ABU08915	U08915 Atropa bell	184	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 112	8	4.0	2086	8	AF224061	AF224061 Solanum a	185	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 113	8	4.0	2086	8	AF224062	AF224062 Solanum a	186	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 114	8	4.0	2086	8	AF224063	AF224063 Normania	187	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 115	8	4.0	2086	8	AF224064	AF224064 Solanum p	188	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 116	8	4.0	2086	8	AF224065	AF224065 Triguera	189	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 117	8	4.0	2086	8	AF224066	AF224066 Solanum v	190	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 118	8	4.0	2086	8	AF224067	AF224067 Solanum e	191	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 119	8	4.0	2086	8	AF224069	AF224069 Solanum m	192	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 120	8	4.0	2086	8	AF224070	AF224070 Solanum v	193	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 121	8	4.0	2086	8	AF224071	AF224071 Solanum c	194	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 122	8	4.0	2086	8	AF224072	AF224072 Solanum c	195	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 123	8	4.0	2086	8	AF224073	AF224073 Solanum j	196	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 124	8	4.0	2086	8	AF224074	AF224074 Solanum m	197	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 125	8	4.0	2086	8	AF224075	AF224075 Solanum m	198	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 126	8	4.0	2086	8	AF224076	AF224076 Solanum m	199	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 127	8	4.0	2086	8	AF224077	AF224077 Solanum m	200	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 128	8	4.0	2086	8	AF224078	AF224078 Solanum m	201	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 129	8	4.0	2086	8	AF224079	AF224079 Solanum m	202	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 130	8	4.0	2086	8	AF224080	AF224080 Solanum m	203	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 131	8	4.0	2086	8	AF224081	AF224081 Solanum m	204	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 132	8	4.0	2086	8	AF224082	AF224082 Solanum m	205	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 133	8	4.0	2086	8	AF224083	AF224083 Solanum m	206	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 134	8	4.0	2086	8	AF224084	AF224084 Solanum m	207	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 135	8	4.0	2086	8	AF224085	AF224085 Solanum m	208	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 136	8	4.0	2086	8	AF224086	AF224086 Solanum m	209	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 137	8	4.0	2086	8	AF224087	AF224087 Solanum m	210	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 138	8	4.0	2086	8	AF224088	AF224088 Solanum m	211	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 139	8	4.0	2086	8	AF224089	AF224089 Solanum m	212	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 140	8	4.0	2086	8	AF224090	AF224090 Solanum m	213	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 141	8	4.0	2086	8	AF224091	AF224091 Solanum m	214	8	4.0	2086	8	WU72755	WU72755 Solanum t
C 142													

C 197	8	4.0	13483	8	AF098806	270	8	4.0	112799	2	AC117860	AC117860	Rattus no
C 198	8	4.0	13782	9	AC108042	271	8	4.0	112910	2	AC131019	AC131019	Rattus no
C 199	8	4.0	13994	1	AE005952	272	8	4.0	113333	2	AC096958	AC096958	Rattus no
C 200	8	4.0	18120	9	HSB20F6	C 273	8	4.0	113759	9	AC008844	AC008844	Homo sapi
C 201	8	4.0	20328	2	AC014603	274	8	4.0	114870	9	AC010755	AC010755	Homo sapi
C 202	8	4.0	25951	5	AB004829	275	8	4.0	115199	9	AC003695	AC003695	Homo sapi
C 203	8	4.0	26590	9	AC090672	C 276	8	4.0	115905	9	HS57E3	HS57E3	Homo sapi
C 204	8	4.0	28170	6	AX283721	C 277	8	4.0	116164	2	AC094266	AC094266	Rattus no
C 205	8	4.0	31664	3	CE110A3	C 278	8	4.0	116356	2	AC094938	AC094938	Rattus no
C 206	8	4.0	32798	6	AX382187	C 279	8	4.0	120558	8	OSJN00125	OSJN00125	Rattus no
C 207	8	4.0	33007	12	AF323988	C 280	8	4.0	120710	8	AP005479	AP005479	Oryza sat
C 208	8	4.0	33592	6	AX084504	C 281	8	4.0	122358	8	AC002396	AC002396	Arabidops
C 209	8	4.0	33699	6	AX084506	C 282	8	4.0	122466	9	HSJ841K13	HSJ841K13	Human DNA
C 210	8	4.0	33988	6	AX084517	C 283	8	4.0	122580	2	AC128369	AC128369	Rattus no
C 211	8	4.0	34303	6	AR091536	C 284	8	4.0	125866	8	OSJN00283	OSJN00283	Oryza sat
C 212	8	4.0	34303	6	AR102229	C 285	8	4.0	126734	8	AP003849	AP003849	Oryza sat
C 213	8	4.0	34341	6	AX084505	C 286	8	4.0	127829	2	AP004154	AP004154	Oryza sat
C 214	8	4.0	34448	6	AX084507	C 287	8	4.0	129311	8	AP003218	AP003218	Oryza sat
C 215	8	4.0	34737	6	AX084518	C 288	8	4.0	129420	8	AC087547	AC087547	Oryza sat
C 216	8	4.0	35408	6	AR163568	C 289	8	4.0	132581	2	AP003742	AP003742	Oryza sat
C 217	8	4.0	35408	6	AR166442	C 290	8	4.0	135469	2	AL136127	AL136127	Homo sapi
C 218	8	4.0	35724	6	AX084516	C 291	8	4.0	136116	2	AC110874	AC110874	Gallus ga
C 219	8	4.0	35764	12	AY046510	C 292	8	4.0	136917	2	AP005097	AP005097	Oryza sat
C 220	8	4.0	35935	6	AR091533	C 293	8	4.0	138589	9	AC092966	AC092966	Homo sapi
C 221	8	4.0	35935	6	AR102226	C 294	8	4.0	138724	2	AC114239	AC114239	Rattus no
C 222	8	4.0	35935	6	AR116313	C 295	8	4.0	139147	2	AC006725	AC006725	Caenorhab
C 223	8	4.0	35935	6	AX451988	C 296	8	4.0	139381	2	OSJN01002	OSJN01002	Oryza sat
C 224	8	4.0	35935	14	ADRCMPGEN	C 297	8	4.0	139480	9	HUAC002331	HUAC002331	Homo sapi
C 225	8	4.0	35937	14	ADRCG	C 298	8	4.0	139646	2	AC115934	AC115934	Mus muscu
C 226	8	4.0	36114	6	AX084519	C 299	8	4.0	140934	2	AC112553	AC112553	Rattus no
C 227	8	4.0	36154	6	AX468857	C 300	8	4.0	141633	8	OSJN00278	OSJN00278	Oryza sat
C 228	8	4.0	36154	6	AX468865	C 301	8	4.0	142689	9	HSB348B13	HSB348B13	Human DNA
C 229	8	4.0	37142	2	AC025045	C 302	8	4.0	142707	2	AP004338	AP004338	Oryza sat
C 230	8	4.0	37391	6	AX468871	C 303	8	4.0	144268	2	OSJN00212	OSJN00212	Oryza sat
C 231	8	4.0	38041	6	AX468869	C 304	8	4.0	144514	2	AC115534	AC115534	Sus scrofa
C 232	8	4.0	38209	1	SC5G9	C 305	8	4.0	144542	9	AC015819	AC015819	Homo sapi
C 233	8	4.0	40897	8	SPCC794	C 306	8	4.0	145467	2	AP004270	AP004270	Oryza sat
C 234	8	4.0	42031	3	CEB35A5	C 307	8	4.0	145668	9	AC092272	AC092272	Homo sapi
C 235	8	4.0	42071	1	SC2G61	C 308	8	4.0	145859	9	AC009510	AC009510	Homo sapi
C 236	8	4.0	46676	9	AL356490	C 309	8	4.0	145887	9	AC091435	AC091435	Homo sapi
C 237	8	4.0	47400	2	AC129582	C 310	8	4.0	146810	2	AL627262	AL627262	Danio rer
C 238	8	4.0	52111	2	AL358796_3	C 311	8	4.0	146921	8	AP002836	AP002836	Oryza sat
C 239	8	4.0	58456	2	AC127514	C 312	8	4.0	147021	9	AC006501	AC006501	Homo sapi
C 240	8	4.0	59000	9	AF314199S1	C 313	8	4.0	147405	2	AC118576	AC118576	Didelphis
C 241	8	4.0	63285	8	AC122785	C 314	8	4.0	150191	8	OSJN00038	OSJN00038	Oryza sat
C 242	8	4.0	63586	8	AC007264	C 315	8	4.0	150626	9	HSJ177P10	HSJ177P10	Human DNA
C 243	8	4.0	71506	9	AC092452	C 316	8	4.0	151559	2	AC015863	AC015863	Homo sapi
C 244	8	4.0	73094	2	AL671114	C 317	8	4.0	151819	2	AC108754	AC108754	Oryza sat
C 245	8	4.0	74774	2	AC100316	C 318	8	4.0	152523	9	AC007009	AC007009	Homo sapi
C 246	8	4.0	75892	2	AC101297	C 319	8	4.0	152699	2	AP005125	AP005125	Oryza sat
C 247	8	4.0	80120	9	HS480J14	C 320	8	4.0	153336	2	AC025020	AC025020	Homo sapi
C 248	8	4.0	81940	9	HS71TIN2B	C 321	8	4.0	153805	9	AC090987	AC090987	Homo sapi
C 249	8	4.0	87551	2	AC021944	C 322	8	4.0	153866	9	AC009307	AC009307	Homo sapi
C 250	8	4.0	88535	2	AC099423	C 323	8	4.0	153926	9	AL160059	AL160059	Human DNA
C 251	8	4.0	89370	3	AC084153	C 324	8	4.0	154123	2	AC019562	AC019562	Drosophil
C 252	8	4.0	89473	8	TIK7	C 325	8	4.0	154153	9	AL445933	AL445933	Human DNA
C 253	8	4.0	91071	8	AC002339	C 326	8	4.0	154210	9	AC078888	AC078888	Homo sapi
C 254	8	4.0	93427	9	AL591076	C 327	8	4.0	154440	2	AC108176	AC108176	Bos tauru
C 255	8	4.0	93475	5	AL672088	C 328	8	4.0	154625	2	AP005383	AP005383	Oryza sat
C 256	8	4.0	94024	9	AC007368	C 329	8	4.0	154683	9	AC006157	AC006157	Homo sapi
C 257	8	4.0	94682	2	AC128917	C 330	8	4.0	154814	9	AC114802	AC114802	Homo sapi
C 258	8	4.0	98980	2	AP004215	C 331	8	4.0	155939	8	CHNTXX	CHNTXX	Nicotiana t
C 259	8	4.0	100000	9	AB020866	C 332	8	4.0	156000	2	AC060815	AC060815	Homo sapi
C 260	8	4.0	100499	9	AL138769	C 333	8	4.0	156277	9	AC107015	AC107015	Homo sapi
C 261	8	4.0	105928	2	AC111925	C 334	8	4.0	156687	8	ABE316582	ABE316582	Atropa be
C 262	8	4.0	107139	2	AL360223	C 335	8	4.0	156856	2	AC097235	AC097235	Rattus no
C 263	8	4.0	107655	2	AC084840	C 336	8	4.0	156929	2	AC026854	AC026854	Homo sapi
C 264	8	4.0	108799	9	AC090989	C 337	8	4.0	157262	9	AC104827	AC104827	Homo sapi
C 265	8	4.0	110000	2	LMFLCHR34_10	C 338	8	4.0	157926	2	AC096047	AC096047	Rattus no
C 266	8	4.0	110000	2	LMFLCHR36_17	C 339	8	4.0	158078	9	AC005094	AC005094	Homo sapi
C 267	8	4.0	110000	2	LMFLCHR36_17	C 340	8	4.0	158133	8	AP002816	AP002816	Oryza sat
C 268	8	4.0	110000	2	AC026388_3	C 341	8	4.0	158242	3	AC115483	AC115483	Drosophil
C 269	8	4.0	110042	9	HS175E3	C 342	8	4.0	158349	9	AC024908	AC024908	Homo sapi

c 343	8	4.0 158719	2	AC112081	AC112081 Rattus no	c 416	8	4.0 183990	8	AF114171	AF114171 Sorghum b
344	8	4.0 159132	2	AC118378	AC118378 Rattus no	c 417	8	4.0 184203	10	AL669913	AL669913 Mouse DNA
345	8	4.0 159694	2	AC098586	AC098586 Homo sapi	c 418	8	4.0 184412	2	AC109147	AC109147 Mus muscu
346	8	4.0 159738	9	AC073352	AC073352 Homo sapi	c 419	8	4.0 185050	2	AC127144	AC127144 Rattus no
347	8	4.0 160557	9	AC117945	AC117945 Homo sapi	c 420	8	4.0 185334	2	AC129392	AC129392 Rattus no
c 348	8	4.0 161247	9	AC069304	AC069304 Homo sapi	421	8	4.0 185635	2	AC122372	AC122372 Mus muscu
349	8	4.0 161437	2	AC093434	AC093434 Homo sapi	422	8	4.0 185741	9	AC006203	AC006203 Homo sapi
c 350	8	4.0 161987	9	AC006070	AC006070 Homo sapi	423	8	4.0 186760	2	AC026056	AC026056 Homo sapi
351	8	4.0 162504	2	AC106209	AC106209 Rattus no	c 424	8	4.0 187046	2	AC079912	AC079912 Mus muscu
c 352	8	4.0 162573	2	AC115535	AC115535 Sus scrof	c 425	8	4.0 187508	9	AC104078	AC104078 Homo sapi
353	8	4.0 162712	2	AC026096	AC026096 Homo sapi	426	8	4.0 187778	2	AC034166	AC034166 Homo sapi
c 354	8	4.0 162860	2	AC119097	AC119097 Rattus no	c 427	8	4.0 187936	9	AC080003	AC080003 Homo sapi
c 355	8	4.0 163282	2	AC016438	AC016438 Homo sapi	428	8	4.0 188324	2	AC0807625	AC0807625 Homo sapi
356	8	4.0 163410	9	AC006198	AC006198 Homo sapi	429	8	4.0 188359	9	AC006369	AC006369 Homo sapi
c 357	8	4.0 163932	2	AC114160	AC114160 Rattus no	c 430	8	4.0 188493	2	AC124181	AC124181 Mus muscu
c 358	8	4.0 164259	2	AC113514	AC113514 Mus muscu	c 431	8	4.0 188507	2	AL844159	AL844159 Mus muscu
c 359	8	4.0 164331	9	AC016254	AC016254 Homo sapi	c 432	8	4.0 189050	1	AL646066	AL646066 Ralstonia
c 360	8	4.0 164958	2	AC092729	AC092729 Canis fam	c 433	8	4.0 189576	2	AL772183	AL772183 Mus muscu
c 361	8	4.0 165228	9	AC005325	AC005325 Homo sapi	434	8	4.0 189838	2	AC090121	AC090121 Mus muscu
c 362	8	4.0 165497	8	AP003232	AP003232 Oryza sat	c 435	8	4.0 190050	1	AL646059	AL646059 Ralstonia
c 363	8	4.0 165866	2	AC087273	AC087273 Homo sapi	c 436	8	4.0 190212	2	AC127097	AC127097 Rattus no
c 364	8	4.0 166474	2	AC107431	AC107431 Rattus no	437	8	4.0 190241	2	AC119764	AC119764 Rattus no
c 365	8	4.0 166517	9	AL356672	AL356672 Human DNA	c 438	8	4.0 190322	9	AC093124	AC093124 Papio cyn
c 366	8	4.0 168662	9	AC079880	AC079880 Homo sapi	c 439	8	4.0 190417	2	AC090319	AC090319 Homo sapi
367	8	4.0 167503	2	AC129344	AC129344 Rattus no	440	8	4.0 190462	9	AC092535	AC092535 Homo sapi
c 368	8	4.0 167627	9	AC079382	AC079382 Homo sapi	441	8	4.0 190624	2	AC013811	AC013811 Homo sapi
369	8	4.0 168367	9	AC104440	AC104440 Homo sapi	c 442	8	4.0 190651	9	AP000781	AP000781 Homo sapi
c 370	8	4.0 168407	2	AC090786	AC090786 Homo sapi	c 443	8	4.0 191098	2	AC073518	AC073518 Homo sapi
c 371	8	4.0 169462	2	AC024033	AC024033 Homo sapi	c 444	8	4.0 191170	9	AC006112	AC006112 Homo sapi
c 372	8	4.0 169747	9	AL355352	AL355352 Human DNA	c 445	8	4.0 191667	2	AC124202	AC124202 Mus muscu
c 373	8	4.0 170711	2	AC032020	AC032020 Homo sapi	446	8	4.0 192126	9	DJ270M14	DJ270M14 Mus muscu
c 374	8	4.0 170761	9	CNS03TER	AL359240 Human chr	447	8	4.0 192130	2	AC116768	AC116768 Mus muscu
c 375	8	4.0 171339	2	AC021566	AC021566 Homo sapi	c 448	8	4.0 192441	2	AC073404	AC073404 Homo sapi
c 376	8	4.0 171676	9	AC107904	AC107904 Homo sapi	449	8	4.0 192809	2	AC110518	AC110518 Mus muscu
c 377	8	4.0 173056	2	AC120613	AC120613 Rattus no	c 450	8	4.0 193495	2	AC108780	AC108780 Mus muscu
c 378	8	4.0 173146	9	AC006370	AC006370 Homo sapi	451	8	4.0 194206	9	AC021863	AC021863 Homo sapi
c 379	8	4.0 173185	2	AC115058	AC115058 Mus muscu	c 452	8	4.0 196606	9	AP000779	AP000779 Homo sapi
380	8	4.0 173239	9	AC016399	AC016399 Homo sapi	453	8	4.0 197176	2	AC022763	AC022763 Homo sapi
381	8	4.0 173422	2	AC117925	AC117925 Rattus no	454	8	4.0 198326	2	AC087839	AC087839 Oryza sat
382	8	4.0 173522	9	AC008756	AC008756 Homo sapi	455	8	4.0 198729	2	AC113292	AC113292 Mus muscu
c 383	8	4.0 174032	2	AC025904	AC025904 Homo sapi	c 456	8	4.0 198748	2	AC1016719	AC1016719 Homo sapi
c 384	8	4.0 174204	2	AC073956	AC073956 Homo sapi	c 457	8	4.0 198929	2	AC101848	AC101848 Mus muscu
c 385	8	4.0 174228	2	AC015105	AC015105 Homo sapi	458	8	4.0 199101	10	AL672034	AL672034 Mouse DNA
c 386	8	4.0 174694	9	AC007217	AC007217 Homo sapi	c 459	8	4.0 199287	2	AC102672	AC102672 Mus muscu
387	8	4.0 175111	2	AC101999	AC101999 Mus muscu	c 460	8	4.0 199534	2	AC013374	AC013374 Homo sapi
c 388	8	4.0 176096	2	AC111117	AC111117 Mus muscu	c 461	8	4.0 199534	2	AC025959	AC025959 Homo sapi
389	8	4.0 176177	2	AC115409	AC115409 Rattus no	c 462	8	4.0 200131	9	AC105213	AC105213 Homo sapi
c 390	8	4.0 176226	2	AC073638	AC073638 Homo sapi	c 463	8	4.0 200240	2	AC011618	AC011618 Homo sapi
c 391	8	4.0 176459	2	AC108684	AC108684 Homo sapi	464	8	4.0 200336	2	AC112955	AC112955 Mus muscu
c 392	8	4.0 176525	2	AL845474	AL845474 Mus muscu	465	8	4.0 201345	2	AL645912	AL645912 Mus muscu
c 393	8	4.0 176624	2	AC104394	AC104394 Homo sapi	466	8	4.0 201780	2	AC124354	AC124354 Mus muscu
394	8	4.0 176841	2	AC117096	AC117096 Rattus no	467	8	4.0 201802	2	AC124473	AC124473 Mus muscu
c 395	8	4.0 177145	2	AC113156	AC113156 Homo sapi	c 468	8	4.0 201982	2	AC027297	AC027297 Mus muscu
c 396	8	4.0 177476	9	AC010680	AC010680 Homo sapi	c 469	8	4.0 202134	2	AL772249	AL772249 Mus muscu
c 397	8	4.0 177954	3	AC011366	AC011366 Homo sapi	c 470	8	4.0 202163	2	AF235099	AF235099 Homo sapi
c 398	8	4.0 178019	3	AC091206	AC091206 Drosophil	c 471	8	4.0 204696	2	AP001654	AP001654 Homo sapi
c 399	8	4.0 178825	2	AP001203	AP001203 Homo sapi	472	8	4.0 205195	2	AC025963	AC025963 Homo sapi
400	8	4.0 179094	2	AP004138	AP004138 Oryza sat	473	8	4.0 205954	9	AC008546	AC008546 Homo sapi
c 401	8	4.0 179161	2	AC122412	AC122412 Mus muscu	c 474	8	4.0 206909	9	AC016397	AC016397 Homo sapi
c 402	8	4.0 179335	3	AP000870	AP000870 Homo sapi	475	8	4.0 209630	2	AC118019	AC118019 Mus muscu
c 403	8	4.0 180000	2	AC004578	AC004578 Homo sapi	c 476	8	4.0 209923	2	AC121806	AC121806 Mus muscu
c 404	8	4.0 180062	2	AC118333	AC118333 Rattus no	c 477	8	4.0 210296	2	AC121971	AC121971 Mus muscu
c 405	8	4.0 180155	2	AL161735	AL161735 Homo sapi	c 478	8	4.0 212080	2	AC091457	AC091457 Mus muscu
c 406	8	4.0 180484	2	AC026759	AC026759 Oryza sat	c 479	8	4.0 213800	2	AC127415	AC127415 Mus muscu
407	8	4.0 180619	9	AL356513	AL356513 Human DNA	c 480	8	4.0 217501	2	AC116128	AC116128 Mus muscu
c 408	8	4.0 180829	2	AC022508	AC022508 Homo sapi	c 481	8	4.0 218648	2	AL731708	AL731708 Mus muscu
c 409	8	4.0 180859	2	AC091126	AC091126 Drosophil	482	8	4.0 219038	2	AC083890	AC083890 Homo sapi
c 410	8	4.0 180968	2	AC078996	AC078996 Mus muscu	c 483	8	4.0 219447	9	HUMFLNGGPD	L44140 Homo sapien
c 411	8	4.0 181415	2	AP005195	AP005195 Oryza sat	c 484	8	4.0 221978	2	AC103664	AC103664 Mus muscu
c 412	8	4.0 181632	2	AC053508	AC053508 Homo sapi	485	8	4.0 222606	2	AC125407	AC125407 Mus muscu
413	8	4.0 181805	10	AL670660	AL670660 Mouse DNA	486	8	4.0 222618	2	AC103335	AC103335 Rattus no
c 414	8	4.0 182440	9	AC096755	AC096755 Homo sapi	487	8	4.0 223717	2	AC108896	AC108896 Bos tauru
c 415	8	4.0 183118	2	AL355528	AL355528 Homo sapi	488	8	4.0 227009	2	AL845264	AL845264 Mus muscu



```

489      8 4.0 228659 2 AC095481 AC095481 Rattus no
490      8 4.0 228856 2 AC087420 AC087420 Mus muscu
491      8 4.0 230851 2 AC115440 AC115440 Rattus no
c 492      8 4.0 231562 9 AC010148 AC010148 Homo sapi
c 493      8 4.0 231772 2 AC125217 AC125217 Mus muscu
c 494      8 4.0 234096 2 AC121818 AC121818 Mus muscu
495      8 4.0 235086 2 AC100043 AC100043 Mus muscu
496      8 4.0 239925 2 AC110921 AC110921 Homo sapi
c 497      8 4.0 247629 2 AC123124 AC123124 Rattus no
c 498      8 4.0 248160 2 AC122000 AC122000 Mus muscu
c 499      8 4.0 257000 1 AP000002 AP000002 Pyrococcu
c 500      8 4.0 257817 2 AC006909 AC006909 Caenorhab

```

## ALIGNMENTS

```

RESULT 1
AF529823
LOCUS      Homo sapiens clone Ramos 9 AID (AID) mRNA, partial cds. linear PRI 19-AUG-2002
DEFINITION
ACCESSION AF529823
VERSION    AF529823.1 GI:22297233
KEYWORDS
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 596)
AUTHORS   Martin,A. and Scharff,M.D.
TITLE     Somatic hypermutation of the AID transgene in B cells and non-B
          cells
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS   Martin,A. and Scharff,M.D.
TITLE     Direct Submission
JOURNAL   Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
          Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
Source     1..596
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="Ramos 9"
            /gene="AID"
            /note="integrated into Burkitt's lymphoma cell line Ramos"
            /codon_start=1
            /product="AID"
            /protein_id="AAM95410.1"
            /db_xref="GI:22297234"
            /translation="MDSLMMNRKFLYOFKNVRWAKGRRETYLCYVVKRRDSATSFSL
            DFGYLRNKGCHVELLRYISDWDLRGRCYRVTWFTSWSPCYDCARHVADFLRGNP
            NLSRIFARLYFCEDRAEPEGLRLHRAVGQIAIMTFKDYFYCWNTFVENHETFK
            AWEGHLSVRLSRLRLLPLLYEVDLRLDAFRTLGL"
BASE COUNT 128 a 163 c 155 g 150 t
ORIGIN

```

## gene

## CDS

```

Alignment Scores:
Pred. No.:      8.06e-201      Length:      596
Score:          198.00        Matches:      198
Percent Similarity: 100.00%    Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match:     100.00%      Indels:      0
DB:              9           Gaps:        0

```

```

US-09-966-880A-8 (1-198) x AF529823 (1-596)
QY      1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
        |||||||
Db       1 ATGGACAGCCYATTGATGACCGGAGGAGATTCTTTACCAATTCAAAAATGTCGCGTGG 60
        |||||||
QY      21 AlaLysGlyArgArgGluThrTyrLeuCystYrValValLysArgAspSerAlaThr 40

```

```

|||||
61 GCTAAGGTCGCGGTGAGACCTACCTGTGCTACGTAGTAGAAGAGCGGTGACAGTCTCTACA 120
QY      41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGlulLeuLeu 60
        |||||||
121 TCCCTTTTTCACGGAGCTTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGAATTGCTC 180
QY      61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
        |||||||
181 TTCTCCCTCATCTCGGACTGGGACCTAGACCTGGCCGCTGCTACCGCGCTCACCCTGG 240
QY      81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
        |||||||
241 TTCACCTCTCGAGCCCTGCTACGACTGTGCCGACATGTGCCGACTTCTTCGCGAGGG 300
QY      101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
        |||||||
301 AACCCCAACCTCAGTCTGAGGATCTCACCGCGCCCTCTACTCTGTGAGAGCCGCAAG 360
QY      121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
        |||||||
361 CTGAGCCCGAGGGCTGCGGGCTGCACCGCGGGGTGCCAANTAGCCATCATGACC 420
QY      141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160
        |||||||
421 TTCAAGATATATTTTACTGCTGGAAATCTTTTGTAGAAACCATGAAAGAACTTTCAAA 480
QY      161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
        |||||||
481 GCCTGGGGAAGGCTGCATGAAATTCAGTCTCTCCAGACACGCTTCGCGCGCATCCT 540
QY      181 LeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
        |||||||
541 TTGCCCTGTATGAGTTGATGACTTACGAGACGCAATTCGTACTTTGGGACTT 594

```

```

RESULT 2
AF529826
LOCUS      Homo sapiens clone Ramos 12 AID (AID) mRNA, partial cds. linear PRI 19-AUG-2002
DEFINITION
ACCESSION AF529826
VERSION    AF529826.1 GI:22297239
KEYWORDS
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 596)
AUTHORS   Martin,A. and Scharff,M.D.
TITLE     Somatic hypermutation of the AID transgene in B cells and non-B
          cells
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS   Martin,A. and Scharff,M.D.
TITLE     Direct Submission
JOURNAL   Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
          Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
Source     1..596
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="Ramos 12"
            /gene="AID"
            /note="integrated into Burkitt's lymphoma cell line Ramos"
            /codon_start=1
            /product="AID"
            /protein_id="AAM95413.1"
            /db_xref="GI:22297240"
            /translation="MDSLMMNRKFLYOFKNVRWAKGRRETYLCYVVKRRDSATSFSL
            DFGYLRNKGCHVELLRYISDWDLRGRCYRVTWFTSWSPCYDCARHVADFLRGNP
            NLSRIFARLYFCEDRAEPEGLRLHRAVGQIAIMTFKDYFYCWNTFVENHETFK
            NLSRIFARLYFCEDRAEPEGLRLHRAVGQIAIMTFKDYFYCWNTFVENHETFK

```

```
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
Source
1..596
/organism.."Mus musculus"
/db_xref.."taxon:10090"
/clone.."10"
/cell_type.."hybridoma P1-5"
/transgenic
1..596
/organism.."Homo sapiens"
/db_xref.."taxon:9606"
1..>596
/gene.."AID"
1..>596
/gene.."AID"
/codon_start..1
/product.."AID"
/protein_id.."AAM95424.1"
/db_xref.."GI:22297262"
/translacion.."MDSLLMNRKFLYQFKNVRWAKGRRETYL..YVVKRRDSATSFSL
DFGLRNKNGCHVELLFLRYISDWDLDPGRCYRVTWFTSPSCYDCARHVAADFLRGNP
NLSRIKFTARLYFCEDRKAPEGLRRLHRAVQIAIMTFKDYFYCWNTFVENHIEKTFK
AMEGLHENSVRSLRQLRILLPLVEVDLDRDAFRTLGL"
BASE COUNT 127 a 163 c 155 g 151 t
ORIGIN
Alignment Scores:
Pred. No.: 8,06e-201 Length: 596
Score: 198.00 Matches: 198
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 12 Gaps: 0
US-09-966-880A-8 (1-198) x AF529837 (1-596)
QY 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
DB 1 ATGGACACCCCTTGATGACCGGAGGAAGTTCTTTACCAATTCAAAATGTCCGCTGG 60
QY 21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40
DB 61 GCTAAGGTCGGCGCTGAGACCTTACCTGTGCTACGTAGTGAAGAGCGGTCAAGTCTACA 120
QY 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
DB 121 TCCCTTTTCACTGGACCTTTGGTTATCTTCGCAATAAGAACGGCTGCCAGTGAATTGCTC 180
QY 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
DB 181 TTCTTCCGCTACATCTCGGACTGGGACCTAGACCTGGCCGCTGCTACCGGTACCTGG 240
QY 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
DB 241 TTCACCTCTCGAGGCCCTGCTACGACTGTGCCGACATGTGGCGGACTTTCTGAGGG 300
QY 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
DB 301 AATCCCAACCTCAGTCTGAGGATCTCACCGCGCCCTCTACTTCTGTGAGGACCGCAAG 360
QY 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
DB 361 GCTGAGCCGAGGGCTCGGGCGGTGCACCGCGCGGGGTGCAANTAGCCATCATGACC 420
QY 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160
DB 421 TTCAAAGATATATTTTACTGCTGGAAATCTTTTGTAGAAAACCATGAAAGAACTTTCAA 480
QY 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
DB 481 GCCTGGGAAGGCTGCATGAATAATTCAGTTCTCTCCAGACACGCTTCGGCGCATCCTT 540
RESULT 3
AF529837
LOCUS Mus musculus clone 10 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION complete cds.
ACCESSION AF529837
VERSION AF529837.1 GI:22297261
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B and non-B cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submission
```

QY 181 LeuProLeuTyrGluValAspLeuArgAspAlaPheArgThrLeuGlyLeu 198  
 |||||  
 Db 541 TTGCCCCGTATGAGTTGATGACTTACGAGACGCATTCGTACTTTGGGACTT 594

RESULT 4  
 BC006296  
 LOCUS 1837 bp mRNA linear PRI 12-JUL-2001  
 DEFINITION Homo sapiens, activation-induced cytidine deaminase, clone  
 MGC:12911 IMAGE:4054915, mRNA, complete cds.  
 ACCESSION BC006296  
 VERSION BC006296.1 GI:13623400  
 KEYWORDS MGC.  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 1837)  
 Strausberg, R.  
 Direct Submission  
 Submitted (09-APR-2001) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 Contact: MGC help desk  
 Email: [cgapbs-re@mail.nih.gov](mailto:cgapbs-re@mail.nih.gov)  
 Tissue Procurement: Louis Staudt  
 cDNA Library Preparation: Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: National Institutes of Health Intramural  
 Sequencing Center (NISC),  
 Gaithersburg, Maryland;  
 Web site: <http://www.nisc.nih.gov/>  
 Contact: [nisc.mgc@nih.gov](mailto:nisc.mgc@nih.gov)  
 Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,  
 Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,  
 Dietrich, N.L., Guan, J., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,  
 Lim, M., Maduro, Q.L., Masiello, C., Mastrian, S.D., McCloskey, J.C.,  
 McDowell, J., Pearson, R., Snyder, B., Stantropop, S., Thomas, P.J.,  
 Tongson, E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A.,  
 Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAL Plate: 17 Row: A Column: 1  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: matched mRNA gi: 9988409.  
 Location/Qualifiers  
 1. .1837  
 /organism="Homo sapiens"  
 /db\_xref="LocusID:57379"  
 /db\_xref="taxon:9606"  
 /clone="MGC:12911 IMAGE:4054915"  
 /tissue-type="Primary B-Cells from Tonsils"  
 /clone\_lib="NIH\_MGC\_48"  
 /lab\_host="DH10B-R"  
 /note="Vector: pOTB7"  
 86. .682  
 /codon\_start=1  
 /product="activation-induced cytidine deaminase"  
 /protein\_id="AAH06296.1"  
 /db\_xref="GI:13623401"  
 /translation="MDSLMMNRKFLYQKNVRWAKGRRETYLCYVVKRRDSATSFSL  
 DFLGRNKGKVELLFYRYSMDWLDPGRCYRVYTWFTWSPCYDCARHVADELRGNP  
 NLSLRIFARLYFCEDKRAEPGLRLRHAGVQIAIMTEKDYFYCWNTFVENHERTEPK  
 AWEGHENSVRLSQLRLPLLYEVDLDRDAFTLGL"  
 530 a 387 c 421 g 499 t

CDS  
 1 (sites)  
 Muto, T., Muramatsu, M., Taniwaki, M., Kinoshita, K. and Honjo, T.  
 Isolation, tissue distribution, and chromosomal localization of the  
 human activation-induced cytidine deaminase (AID) gene  
 Genomics 68 (1), 85-88 (2000)  
 20408890  
 2 (sites)  
 Revy, P., Muto, T., Levy, Y., Geissmann, F., Plebani, A., Sanal, O.,  
 Catalan, N., Forveille, M., Dufourcq-Lagelouse, R., Gennery, A.,  
 Tezcan, I., Ersoy, F., Kayserili, H., Ugazio, A.G., Brousse, N.,  
 Muramatsu, M., Notarangelo, L.D., Kinoshita, K., Honjo, T., Fischer, A.  
 and Durandy, A.  
 Activation-induced cytidine deaminase (AID) deficiency causes the  
 autosomal recessive form of the Hyper-IgM syndrome (HIGM2)  
 Cell 102 (5), 565-575 (2000)

BASE COUNT  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 2,13e-200 Length: 1837  
 Score: 198.00 Matches: 198  
 Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 9 Gaps: 0

US-09-966-880A-8 (1-198) x BC006296 (1-1837)

QY 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20  
 |||||  
 Db 86 ATGGACAGCCCTCTTGTATGAACGGAGGAAGTTCTTTACCAATTCAAAAATGTCCGCTGG 145

QY 21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgSerAlaThr 40  
 |||||  
 Db 146 GCTAAGGTCGGCGTAGACCTACCTGTGTACGTAGTAGAAGAGCGGTGACAGTGTCTACA 205

QY 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60  
 |||||  
 Db 206 TCCCTTTTACATGGACTTTGGTTATCTTCGCAATGAACGGCTGCCAGTGGATTGCTC 265

QY 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80  
 |||||  
 Db 266 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCCGCTGCTACCGGCTCACCTGG 325

QY 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAlaAspPheLeuArgGly 100  
 |||||  
 Db 326 TTCACCTCTCGGAGCCCTGTCTAGCACTGTGCCGACATGTGGCGGACTTCTTCGGAGGG 385

QY 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120  
 |||||  
 Db 386 AACCCCAACCTCAGTCTGAGGATCTTACCCGGCGCCTCTACTTCTGTGAGGAGCGCGAAG 445

QY 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140  
 |||||  
 Db 446 GCTGAGCCGAGGGGCTCGCGGCTGCACCGCGCGGGTGCAAAATAGCCATCATGACC 505

QY 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160  
 |||||  
 Db 506 TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAACCCATGAAGAAGACTTTCAA 565

QY 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180  
 |||||  
 Db 566 GCCTGGGAAGGCTGCATGAAATTCAGTTCTCTCCACACAGCTTCGGGGCATCCTT 625

QY 181 LeuProLeuTyrGluValAspLeuArgAspAlaPheArgThrLeuGlyLeu 198  
 |||||

Db 626 TTGCCCCGTATGAGTTGATGACTTACGAGACGCATTCGTACTTTGGGACTT 679

RESULT 5  
 AB040431  
 LOCUS 2791 bp mRNA linear PRI 03-OCT-2000  
 DEFINITION Homo sapiens AID mRNA for activation-induced cytidine deaminase,  
 complete CDS.  
 ACCESSION AB040431  
 VERSION AB040431.1 GI:9988409  
 KEYWORDS AID: activation-induced cytidine deaminase; Human AID.  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (sites)  
 Muto, T., Muramatsu, M., Taniwaki, M., Kinoshita, K. and Honjo, T.  
 Isolation, tissue distribution, and chromosomal localization of the  
 human activation-induced cytidine deaminase (AID) gene  
 Genomics 68 (1), 85-88 (2000)  
 20408890  
 2 (sites)  
 Revy, P., Muto, T., Levy, Y., Geissmann, F., Plebani, A., Sanal, O.,  
 Catalan, N., Forveille, M., Dufourcq-Lagelouse, R., Gennery, A.,  
 Tezcan, I., Ersoy, F., Kayserili, H., Ugazio, A.G., Brousse, N.,  
 Muramatsu, M., Notarangelo, L.D., Kinoshita, K., Honjo, T., Fischer, A.  
 and Durandy, A.  
 Activation-induced cytidine deaminase (AID) deficiency causes the  
 autosomal recessive form of the Hyper-IgM syndrome (HIGM2)  
 Cell 102 (5), 565-575 (2000)

TITLE  
 JOURNAL

```
MEDLINE 20460541
REFERENCE 3 (bases 1 to 2791)
AUTHORS Muto,T., Muramatsu,M., Taniwaki,M., Kinoshita,K. and Honjo,T.
TITLE Direct Submission
JOURNAL Submitted (18-MAR-2000) Tasuku Honjo, Kyoto University, Department
of Medical Chemistry, Faculty of Medicine; Yoshida, Sakyo-ku,
Kyoto, Kyoto 606-8501, Japan (E-mail:honjo@mfour.med.kyoto-u.ac.jp,
Tel:81-75-753-4371(ex.4371), Fax:81-75-753-4388)
FEATURES
    source
    Location/Qualifiers
        1..2791
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
        1..2791
            /gene="AID"
        77..673
            /gene="AID"
        /codon_start=1
        /product="activation-induced cytidine deaminase"
        /protein_id="BAB12721.1"
        /db_xref="GI:9988410"
        /translation="MDSLMMNRKFLYQFKNVRKGRRETYLCYVVKRRDSATSFSL
        DFGYLRNKGCHVELLFLRYISDMDLDGRCYRVWTFSWSPCYDCARHVDPLRGNP
        NLSRIETARLYFCEDRKAEPEGLRLHRAGVQIAIMTFKDYFCWNTFVENHERTEK
        AWELGHSNLSRLSQLRILLPLLYEVDLDRDAFTLGL"
BASE COUNT 842 a 548 c 625 g 776 t
ORIGIN
Alignment Scores:
Pred. No.: 3,066-200 Length: 2791
Score: 198.00 Matches: 198
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-09-966-880A-8 (1-198) x AB040431 (1-2791)

Qy 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrrp 20
Db 77 ATGGACACCCCTTGATGAACGGGAGGAGTTCTTTACCAATTCAAAATATGTCGGCTGG 136

Qy 21 AlaLysGlyArgArgGluThrTyrLeuCystTyrValValLysArgArgAspSerAlaThr 40
Db 137 GCTAAGGTCGCGCTGAGACCTACCTGTGTACGTAGTGTAGGAGCGGTGACAGTGCTACA 196

Qy 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
Db 197 TCCITTTTCACGTGGACTTTGGTTATCTTCGCAATAGAACGCGCTGCCAGTGGAAATGGCTC 256

Qy 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
Db 257 TTCTCTCCGCTACATCTCGGACTGGGACCTAGACCTGCGCGCTGCTACCGCGTCACTGG 316

Qy 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
Db 317 TTCACCTCTGGAGCCCTGCTAGACATGTGCCGACATGTGGCGGACTTTCTGCGAGGG 376

Qy 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
Db 377 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGGCTCTACTCTGTGAGGACCGCAAG 436

Qy 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
Db 437 GCTGAGCCGAGGGGCTCGCGGCGCTGCACCGCGCGGGTGCAAAATAGCCATCATGACC 496

Qy 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160
Db 497 TTCAAAGATTATTTTACTGCTGGAATACTTTGTAGAAAACCATGAAGAATTTTCAA 556

Qy 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
Db 557 GCCTGGGAAGGCGTGCATGAAATTCAGTTCGTCTCTCCAGACAGCTTCGCGGCATCCTT 616
```

```
Qy 181 LeuProLeuTyrGluValAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
Db 617 TTGCCCCCTGTATGAGGTGATGACTTACGAGACCCATTTCTGCTACTTTGGGACTT 670

RESULT 6
AF529829
LOCUS Mus musculus clone 2 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION complete cds.
ACCESSION AF529829
VERSION AF529829.1 GI:22297245
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B and non-B cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
    source
    Location/Qualifiers
        1..596
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /clone="2"
            /cell_type="hybridoma PI-5"
            /transgenic
        1..596
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
        1..596
            /gene="AID"
        1..596
            /gene="AID"
        /codon_start=1
        /product="AID"
        /protein_id="AAM95416.1"
        /db_xref="GI:22297246"
        /translation="MDRLMMNRKFLYQFKNVRKGRRETYLCYVVKRRDSATSFSL
        DFGYLRNKGCHVELLFLRYISDMDLDGRCYRVWTFSWSPCYDCARHVDPLRGNP
        NLSRIETARLYFCEDRKAEPEGLRLHRAGVQIAIMTFKDYFCWNTFVENHERTEK
        AWELGHSNLSRLSQLRILLPLLYEVDLDRDAFTLGL"
BASE COUNT 128 a 163 c 155 g 150 t
ORIGIN
Alignment Scores:
Pred. No.: 1,246-197 Length: 596
Score: 195.00 Matches: 195
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 98.48% Indels: 0
DB: 12 Gaps: 0

US-09-966-880A-8 (1-198) x AF529829 (1-596)

Qy 4 LeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrrpAlaLysGly 23
Db 10 CTCTTTGATGAACCGGAGGAGTTCTTTTACCAATTCAAAATATGTCGGCTGACGGT 69

Qy 24 ArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThrSerPheSer 43
Db 70 CGGCGTGAGACCTACCTGTGTAGTGTAGTGAAGAGCGGTGACAGTGCTACATCCTTTTCA 129

Qy 44 LeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeuPheLeuArg 63
Db 130 CTGACCTTTGGTTATCTTCGCAATAGAACGCGTGCACCGTGGAAATGCTCTCTCCGCG 189

Qy 64 TyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrpPheThrSer 83
```

```
Db      190 TACATCTCGGACTGGAGCTAGACCTGGCGCTGCTACCGCGTCACTGTTCACTCC 249
Qy      84 TrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGlyAsnProAsn 103
Db      250 TGGAGCCCTGCTAGGACTGTGCCGACATGTGCCGACTTTCGCGAGGAACCCCAAC 309
Qy      104 LeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgGlyAlaGluPro 123
Db      310 CTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAGCTGAGGCC 369
Qy      124 GluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThrPheLysAsp 143
Db      370 GAGGGCGTCCGGCGCTGCACCGCGCGGTGCAAAATAGCATCATGACCTTCAAGAT 429
Qy      144 TyrPheTyrCysTyrAsnThrPheValGluAsnHisGluArgThrPheLysAlaTrpGlu 163
Db      430 TATTTTACTGCTGGAATACCTTTGTAGAAAACCATGAAGAAGCTTCAAGACCTGGAA 489
Qy      164 GlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeuLeuProLeu 183
Db      490 GGGCTGCATGAAATTCAGTTCTGCTCTCCACAGACGCTTCGGCGCATCTTTTGGCCCTG 549
Qy      184 TyrGluValAspAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
Db      550 TATGAGGTTGATGACTTACGAGACGCAATTCGTACTTTGGGACTT 594

RESULT 7
AF529842 597 bp mRNA linear SYN 19-AUG-2002
LOCUS Cricetulus griseus clone 2 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION complete cds.
ACCESSION AF529842
VERSION AF529842.1 GI:22297271
KEYWORDS Chinese hamster.
SOURCE Cricetulus griseus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
Cricetulus.
REFERENCE 1 (bases 1 to 597)
AUTHORS Martin, A. and Scharff, M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 597)
AUTHORS Martin, A. and Scharff, M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source
1..597
/organism="Cricetulus griseus"
/db_xref="taxon:10029"
/clone="2"
/cell_line="CHO"
/transgenic
1..597
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..597
/gene="AID"
1..597
/gene="AID"
/codon_start=1
/product="AID"
/protein_id="AA095429.1"
/db_xref="GI:22297272"
/translation="MDSLMMNRKFLYQKNVRWAKGRRETYLCYVVKRRDSATFSFL
DFGLRNKGCHVELLEFLRYSIDWLDPGRCYRVTFWFTSNPSPCYADHADFRLGNP
NLSLRIFATRYFCEDRKAEPGLRLHRAGVQIAIMTFKDYFCWNTFVENHETFK
AWEGHNSVRLSRQLRILLPLLYEVDLDRAFTWGR"

BASE COUNT 127 a 164 c 157 g 149 t
```

```
ORIGIN
Alignment Scores: 1.25e-197 Length: 597
Pred. No.: 195.00 Matches: 195
Score: 100.00% Conservative: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 98.48%
DB: 12 Gaps: 0

US-09-966-880a-8 (1-198) x AF529842 (1-597)
Qy      1 MetAspSerLeuLeuMetAsnArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
Db      1 ATGACACGCTCTTTCATGACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGCTGG 60
Qy      21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40
Db      61 GCTAAGGGTCGGCGTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGAGAGTCTACA 120
Qy      41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
Db      121 TCCCTTTTCACGTGGACTTTGGTTATCTTCGCAATGAAGACGGCTGCCACGTGGAATTGCTC 180
Qy      61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
Db      181 TTCCTCCGCTACATCTCGGACTTGGGACCTAGACCTGGCCGCTGCTACCCGCTCACCTGG 240
Qy      81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
Db      241 TTCACCTCTCTGGAGCCCTGCTACGACTGTCCCGACATGTGGCGGACTTTCTGCGAGGG 300
Qy      101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
Db      301 AACCCCAACCTCAGCTCTGAGGATCTTCAACCGCGCGCTCTACTTCTGTGAGACCGCAAG 360
Qy      121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
Db      361 GCTGAGCCCGAGGGGCTGCGCGGCTGCACCGCGCGGGTGCAAATAGCCATCATGACC 420
Qy      141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160
Db      421 TTCAAGAGATATATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAAGAACTTTCAA 480
Qy      161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
Db      481 GCCTGGGAAGGCGTCATGAAATTCAGTTCTCTCCACACAGCTTCGGCGGCATCCTT 540
Qy      181 LeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThr 195
Db      541 TTGCCCTGTATGAGGTTGATGACTTACGTGACGCGCATTTTCGTACT 585

RESULT 8
AF529843 597 bp mRNA linear SYN 19-AUG-2002
LOCUS Cricetulus griseus clone 3 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION complete cds.
ACCESSION AF529843
VERSION AF529843.1 GI:22297273
KEYWORDS Chinese hamster.
SOURCE Cricetulus griseus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
Cricetulus.
REFERENCE 1 (bases 1 to 597)
AUTHORS Martin, A. and Scharff, M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 597)
AUTHORS Martin, A. and Scharff, M.D.
TITLE Direct Submission
```

JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA

FEATURES  
source 1..597  
/organism="Cricetulus griseus"  
/db\_xref="taxon:10029"  
/clone="3"  
/cell\_line="CHO"  
/transgenic  
source 1..597  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
1..597  
/gene="AID"  
1..597  
/gene="AID"  
/codon\_start=1  
/product="AID"  
/protein\_id="AAM95430.1"  
/db\_xref="GI:22297274"  
/translation="MDSLMMNRKFLYQFNVRWAKGRRETYLCYVVKRDSATSFSL  
DFGLRNKNGCHVELLRLYISDWLDPGRCYRVWTISWSPCYDCARHVDLFRGNP  
NLSRIPTARLYFCEDRKAPGELRLRHAGVOIAIMTFKDYFCWMTFVENHRTFK  
AWEGLHNSVRLSQLRILLPLYEVDLDRDAFTWGR"  
BASE COUNT 129 a 164 c 156 g 148 t  
ORIGIN

Alignment Scores:  
Pred. No.: 1.25e-197 Length: 597  
Score: 195.00 Matches: 195  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 98.48% Indels: 0  
DB: 12 Gaps: 0

US-09-966-880A-8 (1-198) x AF529843 (1-597)

Qy 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20  
Db 1 ATGCACAGCCTCTTGATGAACCGGAGGAGTTCTTTACCAATTCAAAATGTCGCGTGG 60

Qy 21 AlalysGlyArgArgGluThrTyrLeuCystyrValValLysArgArgAspSerAlaThr 40  
Db 61 GCTAAGGTCGGCGTGAGACCTACTCTGTCTAGCTAGTGAAGAGCGCTGACAGTCTACA 120

Qy 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60  
Db 121 TCCTTTTCACTGGACTTTGGTTATCTTCGCAATGAAGACCGCTCCACGTTGGAATTGCTC 180

Qy 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80  
Db 181 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGCGCGCTGCTACCGCGTCAACCTGG 240

Qy 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100  
Db 241 TTCACCTCTCGAGCCCTGCTAGACTGTGCGCGACATGTGGCGGACTTTCTCGGAGGG 300

Qy 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120  
Db 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360

Qy 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140  
Db 361 GCTGAGCCCGAGGGGCTGCGCGGCTGCACCGCGCGGGTGCAATAGCATCATGACC 420

Qy 141 PheLysAspTyrPheTyrCysTyrAsnThrPheValGluAsnHisGluArgThrPheLys 160  
Db 421 TTCAAGATTATTTTACTGCTGGAATACITTTGTAGAAAAACCATGAAGAACTTTCAAA 480

Qy 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180  
Db 481 GCCTGGGAAGGGCTACATGAAATTCAGTTGCTCTCCAGACAGCTTCGCGGCATCCTT 540

Qy 181 LeuProLeuTyrGluValAspLeuArgAspAlaPheArgThr 195  
Db 541 TTGCCCTGTATGAGGTTGATGACTTAGGAGACGCATTTCTGTACT 585

RESULT 9  
AF529844 597 bp mRNA linear SYN 19-AUG-2002  
LOCUS Cricetulus griseus clone 4 transgenic Homo sapiens AID (AID) mRNA,  
DEFINITION complete cds.  
ACCESSION AF529844  
VERSION AF529844.1 GI:22297275  
KEYWORDS Chinese hamster.  
SOURCE Cricetulus griseus  
ORGANISM Cricetulus griseus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
Cricetulus.  
REFERENCE 1 (bases 1 to 597)  
AUTHORS Martin,A. and Scharff,M.D.  
TITLE Somatic hypermutation of the AID transgene in B cells and non-B  
cells  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 597)  
AUTHORS Martin,A. and Scharff,M.D.  
TITLE Direct Submission  
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of  
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA  
FEATURES  
source 1..597  
/organism="Cricetulus griseus"  
/db\_xref="taxon:10029"  
/clone="4"  
/cell\_line="CHO"  
/transgenic  
1..597  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
1..597  
/gene="AID"  
1..597  
/gene="AID"  
/codon\_start=1  
/product="AID"  
/protein\_id="AAM95431.1"  
/db\_xref="GI:22297276"  
/translation="MDSLMMNRKFLYQFNVRWAKGRRETYLCYVVKRDSATSFSL  
DFGLRNKNGCHVELLRLYISDWLDPGRCYRVWTISWSPCYDCARHVDLFRGNP  
NLSRIPTARLYFCEDRKAPGELRLRHAGVOIAIMTFKDYFCWMTFVENHRTFK  
AWEGLHNSVRLSQLRILLPLYEVDLDRDAFTWGR"  
BASE COUNT 128 a 163 c 158 g 148 t  
ORIGIN

Alignment Scores:  
Pred. No.: 1.25e-197 Length: 597  
Score: 195.00 Matches: 195  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 98.48% Indels: 0  
DB: 12 Gaps: 0

US-09-966-880A-8 (1-198) x AF529844 (1-597)

Qy 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20  
Db 1 ATGCACAGCCTCTTGATGAACCGGAGGAGTTCTTTACCAATTCAAAATGTCGCGTGG 60

Qy 21 AlalysGlyArgArgGluThrTyrLeuCystyrValValLysArgArgAspSerAlaThr 40  
Db 61 GCTAAGGTCGGCGTGAGACCTACTCTGTCTAGCTAGTGAAGAGCGCTGACAGTCTACA 120

Qy 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60  
Db 121 TCCTTTTCACTGGACTTTGGTTATCTTCGCAATGAAGACCGCTCCACGTTGGAATTGCTC 180

```
QY 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
|||||
Db 181 TTCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCGCTGCTACCGCTCACCTGG 240

QY 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
|||||
Db 241 TTCACTCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACTTCTTGGGAGG 300

QY 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
|||||
Db 301 AACCCCAACTCAGTCTGAGGATCTACCGCGCGCTCTACTTCTGTGAGGACCGGAG 360

QY 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
|||||
Db 361 GCTGAGCCGAGGGCTGGCGGCTGCACCGCGGGGTGCAAAATAGCCATCATGACC 420

QY 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160
|||||
Db 421 TTCAAAGATATATTTTACTGCTGGAATACTTTGTAGAAAACCATGAAGAATTTCAA 480

QY 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
|||||
Db 481 GCCTGGGAGGGCTGCATGAATAATTCAGTTCTCTCTCCAGACAGCTTCGGCGCATCCTT 540

QY 181 LeuProLeuTyrGluValAspLeuArgAspAlaPheArgThr 195
|||||
Db 541 TTGCCCTCTATGAGGTTGATGACTTACGAGACGCATTTTCGTACT 585

RESULT 10
AF529847 597 bp mRNA linear SYN 19-AUG-2002
LOCUS Cricetus griseus clone 7 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION complete cds.
ACCESSION AF529847
VERSION AF529847.1 GI:22297281
KEYWORDS Chinese hamster.
SOURCE Cricetus griseus
ORGANISM Cricetus griseus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
Cricetus
REFERENCE 1 (bases 1 to 597)
AUTHORS Martin.A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL Unpublished
AUTHORS Martin.A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
Source 1..597
/organism="Cricetus griseus"
/db_xref="taxon:10029"
/clone="7"
/cell_line="CHO"
/transgenic
1..597
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..597
/gene="AID"
1..597
/gene="AID"
/codon_start=1
/product="AID"
/protein_id="AA095434.1"
/db_xref="GI:22297282"
/translation="WDSLMMNRKFLYOFKNVRWAKGRRETYLCYVVKRDSATSFSL
DFGYLRNKGCHVELLFYRISDWLDGRGYRVKVTWFTSWSPCYCARHVDFLRGNP
NLSLRIFTARLFCEDRKAEPGLRLRHAGVQIAIMTFKDYFCWNTFVENHRTFK
```

```
BASE COUNT 129 a 164 c 156 g 148 t
ORIGIN
AWEGLEHNSVRLSRQLRRILLPLYEVDLDRDAFTWGR"
Alignment Scores: 1.25e-197 Length: 597
Pred. No.: 195.00 Matches: 195
Score: 100.00% Conservative: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 98.48% Gaps: 0
DB: 12
US-09-966-880A-8 (1-198) x AF529847 (1-597)

QY 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTTrp 20
|||||
Db 1 ATGGACAGCCCTCTTGATGACCGGAGGAATTTCTTTACCAATTCAAAAATGTCCGCTGG 60

QY 21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40
|||||
Db 61 GCTAAGGGTGGCGGTGAGACCTACCTGTGCTACGTAGTAGTGAAGAGCGGTGACAGTGCTACA 120

QY 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
|||||
Db 121 TCCTTTTTCACCTGGAGCTTTGGTTATCTTCGCAATAAGACGGCTGCCACGTGGAATTTGCTC 180

QY 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
|||||
Db 181 TTCTCCGCTACATCTCGGACTGGGACCTAGACCTGTGCCGACTGTGCCGACTTCTTTCGCGAGG 240

QY 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
|||||
Db 241 TTCACTCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACTGTGCCGACTTCTTTCGCGAGG 300

QY 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
|||||
Db 301 AACCCCAACTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGGAG 360

QY 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
|||||
Db 361 GCTGAGCCGAGGGCTGGCGGCTGCACCGCGGGGTGCAAAATAGCCATCATGACC 420

QY 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160
|||||
Db 421 TTCAAAGATATATTTTACTGCTGGAATACTTTGTAGAAAACCATGAAGAATTTCAA 480

QY 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
|||||
Db 481 GCCTGGGAGGGCTGCATGAATAATTCAGTTCTCTCTCCAGACAGCTTTCGGCGCATCCTT 540

QY 181 LeuProLeuTyrGluValAspLeuArgAspAlaPheArgThr 195
|||||
Db 541 TTGCCCTCTATGAGGTTGATGACTTACGAGACGCATTTTCGTACT 585

RESULT 11
AF529852 597 bp mRNA linear SYN 19-AUG-2002
LOCUS Cricetus griseus clone 12 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION complete cds.
ACCESSION AF529852
VERSION AF529852.1 GI:22297291
KEYWORDS Chinese hamster.
SOURCE Cricetus griseus
ORGANISM Cricetus griseus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
Cricetus
REFERENCE 1 (bases 1 to 597)
AUTHORS Martin.A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 597)
```

AUTHORS Martin.A. and Scharff,M.D.  
 TITLE Direct Submission  
 JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA

FEATURES  
 source 1..597  
 /organism="Cricetulus griseus"  
 /db\_xref="taxon:10029"  
 /clone="12"  
 /cell\_line="CHO"  
 /transgenic  
 source 1..597  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 gene 1..597  
 /gene="AID"  
 CDS 1..597  
 /product="AID"  
 /protein\_id="AAM95439.1"  
 /db\_xref="GI:22297292"  
 /translation="MDSLNNRRKFLYQKFNVRWAKGRRETYLYVVKRRDSATSFSLDFGLRNKNGCHVELLFURYISDWLDPRGRCYRVTFWFTSWSPCYDCARHVAFLRGNPNLSLRIFTARLYFCEDRKAEPGLRLRHAGVQIAIMTFKDYFCWNTFVENHRTFKAWEGHLSVRLSRLRILLPLYEVDDLRFDAFTWGR"

BASE COUNT 129 a 164 c 156 g 148 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 1,25e-197 Length: 597  
 Score: 195.00 Matches: 195  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 98.48% Indels: 0  
 DB: 12 Gaps: 0

US-09-966-880A-8 (1-198) x AF529852 (1-597)

QY 1 MetAspSerLeuLeuMetAsnArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20  
 Db 1 ATGGACAGCCTCTTGATGAACGGAGGAAGTTCTTTACCAATTCAAAATTCGCGCTGG 60  
 QY 21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40  
 Db 61 GCTAAGGTCGCGCTGAGACCTACCTGTCTAGTAGTGAAGAGCGGTGACAGTGCTACA 120  
 QY 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60  
 Db 121 TCCTTTTCACTGACCTTTGGTTATCTTCGAATAAGACGGCTGCCACGTGCAATTGCTC 180  
 QY 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80  
 Db 181 TTCCCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCCGCTGCTACCGCTACCTGG 240  
 QY 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100  
 Db 241 TTCACCTCTGAGCCCTGCTACGACTGTGCCGACATGTGCCGCACTTCTGCGAGGG 300  
 QY 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120  
 Db 301 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360  
 QY 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140  
 Db 361 GCTGAGCCGAGGAGCTGGCGCGCTGGACCGCGGGGTGCAATAGCCATCATGACC 420  
 QY 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160  
 Db 421 TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAGAATCTTCAAA 480  
 QY 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180

Db 481 GCCTGGGAAGGCTGCATGAAATTCAGTTCGTCTCTCCAGACAGCTTCGCGGCATCCCT 540  
 QY 181 LeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThr 195  
 Db 541 TTGCCCCCTGATGAGGTTGATGACATTACGAGACGCAATTCGTACT 585

RESULT 12  
 AF529846  
 LOCUS Cricetulus griseus clone 5 transgenic Homo sapiens AID (AID) mRNA, complete cds. 597 bp mRNA linear SYN 19-AUG-2002  
 DEFINITION AF529846  
 VERSION AF529846  
 KEYWORDS AF529846.1 GI:22297279  
 SOURCE Chinese hamster.  
 ORGANISM Cricetulus griseus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetulus.  
 REFERENCE 1 (bases 1 to 597)  
 AUTHORS Martin,A. and Scharff,M.D.  
 TITLE Somatic hypermutation of the AID transgene in B cells and non-B cells  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 597)  
 AUTHORS Martin,A. and Scharff,M.D.  
 TITLE Direct Submission  
 JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA

FEATURES  
 source 1..597  
 /organism="Cricetulus griseus"  
 /db\_xref="taxon:10029"  
 /clone="16"  
 /cell\_line="CHO"  
 /transgenic  
 source 1..597  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 gene 1..597  
 /gene="AID"  
 CDS 1..597  
 /gene="AID"  
 /codon\_start=1  
 /product="AID"  
 /protein\_id="AAM95433.1"  
 /db\_xref="GI:22297280"  
 /translation="MDSLNNRRKFLYQKFNVRWAKGRRETYLYVVKRRDSATSFSLDFGLRNKNGCHVELLFURYISDWLDPRGRCYRVTFWFTSWSPCYDCARHVAFLRGNPNLSLRIFTARLYFCEDRKAEPGLRLRHAGVQIAIMTFKDYFCWNTFVENHRTFKAWEGHLSVRLSRLRILLPLYEVDDLRFDAFTWGR"

BASE COUNT 128 a 165 c 157 g 147 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 2,57e-192 Length: 597  
 Score: 190.00 Matches: 190  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 95.96% Indels: 0  
 DB: 12 Gaps: 0

US-09-966-880A-8 (1-198) x AF529846 (1-597)

QY 6 MetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrpAlaLysGlyArgArg 25  
 Db 16 ATGAACCGAGGAAGTTCTTTACCAATTCAAAATTCGCGCTAAGCGTCGCGT 75  
 QY 26 GluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThrSerPheSerLeuAsp 45  
 Db 76 GAGACCTACCTGCTGCTACGTAGTGAAGAGCGGTGACGTGCTACATCTTTTCATCGGAC 135  
 QY 46 PheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeuPheLeuArgIle 65



```
Db 136 TTTGGTTATCTCGCAATAAGACGGCTGCCACGTGGAAATTCGCTTCTCCGCTACATC 195
Qy 66 SerAspTtpAspLeuAspProGlyArgCysTyrArgValThrTrpPheThrSerTrpSer 85
Db 196 TCGACTGGACCTAGACCCCTGGCGCGTCTACCGCGTCACTGTTCCACCTCTCTGGAGC 255
Qy 86 ProCystyrAspCysAlaArgHisValAlaAspPheLeuArgGlyAsnProAsnLeuSer 105
Db 256 CCCTGCTACGACTGTGCCGACATGTGGCCGACTTCTTCTGCGAGGGAACCCCAACCTCAGT 315
Qy 106 LeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLysAlaGluProGluGly 125
Db 316 CTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCAAGGCTGAGCCCGAGGGG 375
Qy 126 LeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThrPheLysAspTyrPhe 145
Db 376 CTGGCGCGCTGCACCGCGCGGGGTGCAATAGCCATCATGACCTTCAAGATTAATTTT 435
Qy 146 TyrCysTtpAsnThrPheValIgluAsnHisGluArgThrPheLysAlaTtpGluGlyLeu 165
Db 436 TACTGCTGGAATCTTTTGTAGAAAACCATGAAAGAACTTTCAAAGCCTGGGAAGGGCTG 495
Qy 166 HisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeuLeuProLeuTyrGlu 185
Db 496 CATGAAAATTCAGTTCGCTCTCCAGACAGCTTCGGCGCATCTTTTGGCCCTGTATGAG 555
Qy 186 ValAspAspLeuArgAspAlaPheArgThr 195
Db 556 GTTGATGACTTACGAGACGCATTTTCGTACT 585

RESULT 13
AF529832 596 bp mRNA linear SYN 19-AUG-2002
LOCUS Mus musculus clone 5 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION complete cds.
ACCESSION AF529832
VERSION AF529832.1 GI:22297251
KEYWORDS house mouse.
SOURCE Mus musculus.
ORGANISM Mus musculus.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 596)
JOURNAL Martin,A. and Scharff,M.D.
REFERENCE Somatic hypermutation of the AID transgene in B and non-B cells
AUTHORS 2 (bases 1 to 596)
TITLE Martin,A. and Scharff,M.D.
JOURNAL Direct Submission
FEATURES
SOURCE Location/Qualifiers
1..596
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="5"
/cell_type="hybridoma P1-5"
/transgenic
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..>596
/gene="AID"
1..>596
/gene="AID"
/codon_start=1
/product="AID"
1..>596
/db_xref="GI:22297252"
/translation="MDSLMMRRKFLYOFKNVRWAKGRRETYLCYVVKRRDPSFSL
DPGYLRNKGCHVELLFELYSIDNDLDFGRCYRVWTWSPCDCAKHVADFLRGNP
NLSLRIFTARDYFCEDRKAEPEGLRLRHAGVQIAIMTFKDYFCWNTFVENHETFK
```

```
BASE COUNT 129 a 164 c 153 g 150 t
ORIGIN
AWEGLHENSRLSRQLRHILLPLYEVDLDAFRTLGL"
Alignment Scores: 1.68e-178 Length: 596
Pred. No.: 177.00 Matches: 177
Score: 100.00% Conservative: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 89.39% Indels: 0
Query Match: 12 Gaps: 0
DB: 0

US-09-966-880A-8 (1-198) x AF529832 (1-596)
Qy 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
Db 1 ATGCACAGCCCTCTTGATGAAACCGGAGGAAGTTCTTTTACCANTTCAAAAATGTCGCTGG 60
Qy 21 AlalysGlyArgArGgluThrTyrLeuCysTyrValValLysArGArgAspSerAlaThr 40
Db 61 GCTAAAGGTGCGCGTGAACCTACCTGTGCTACGTAGTGAAGAGGCGTGACAGTGCTACA 120
Qy 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValIgluLeu 60
Db 121 TCCTTTTTCACGTGACTTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGAATTC 180
Qy 61 PheLeuArgTyrIleSerAspTtpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
Db 181 TTCCTCCGTACATCTCGGACTGGGACCTAGACCCCTGGCCCGCTGCTACCGCTCACCTGG 240
Qy 81 PheThrSerTtpSerProCystyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
Db 241 TTCACCTCTCTGAGCCCTCTACGACTGTGCCGACATGTGGCCGACTTTCTGCGAGGG 300
Qy 101 AsnProAsnLeuSerLeuArgIlePhePheThrIleArgLeuTyrPheCysGluAspArgLys 120
Db 301 AACCCCAACCTCAGTCTGAGGATCTTACCCCGCGCCCTCTACTTCTGTGAGGACCGCAAG 360
Qy 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
Db 361 GCTGAGCCCGAGGGGCTCGCGCGCTGCACCGCGCGGGGTGCAAAATAGCCATCATGACC 420
Qy 141 PheLysAspTyrPheTyrCysTtpAsnThrPheValGluAsnHisGluArgThrPheLys 160
Db 421 TTCAAAGATATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAAGAACTTCAAA 480
Qy 161 AlaTtpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArg 177
Db 481 GCCTGGGAAGGGCTGCATGATAAAATTCAGTTCGTCTCTCCACACAGCTTTCGG 531

RESULT 14
AF529815 596 bp mRNA linear PRI 19-AUG-2002
LOCUS Homo sapiens clone Ramos 1 AID (AID) mRNA, partial cds.
DEFINITION AF529815
ACCESSION AF529815.1 GI:22297217
VERSION AF529815.1
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 596)
JOURNAL Martin,A. and Scharff,M.D.
REFERENCE Somatic hypermutation of the AID transgene in B cells and non-B
AUTHORS cells
TITLE Unpublished
JOURNAL 2 (bases 1 to 596)
REFERENCE Martin,A. and Scharff,M.D.
AUTHORS Direct Submission
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES Location/Qualifiers
1..596
source
```

```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Ramos 1"
1..>596
/gene="AID"
1..>596
/gene="AID"
/notes="integrated into Burkitt's lymphoma cell line Ramos"
/codon_start=1
/product="AID"
/protein_id="AAM95402.1"
/db_xref="GI:22297218"
/translacion="MDSLMMNRKFLYQFNVRWAKGRRETYLCYVVKRRDSATSFSL
DFGYLRNKNCHVELLEFLRYISDWLDPGRCYRVTFWTSRSPCYDCAHVADELGRNP
NLSLRIFATRLYFCEDRKAEPEGLRLHRAGVQIAIMTFKDYFCWMTFVENHERTFK
AWEGLEHNSVRLSRQIRILLPLVEVDLDRDAFRTLGL"
BASE COUNT 128 a 163 c 155 g 150 t
ORIGIN

Alignment Scores:
Pred. No.: 2,24e-176 Length: 596
Score: 175.00 Matches: 175
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 88.38% Indels: 0
DB: 9 Gaps: 0

US-09-966-880A-8 (1-198) x AF529815 (1-596)
Qy 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
Db 1 ATGCACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGCTGG 60
Qy 21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40
Db 61 GCTAAGGTCGGCTGAGACCTACTCTGTCTAGTGTGAAGAGCGCTGACAGTGTCTACA 120
Qy 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
Db 121 TCCTTTTCACTGGACTTTGGTTATCTTCGCANTAAGACGCTGCCAGTGGAAATGCTC 180
Qy 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
Db 181 TTCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCCGCTGCTACCGCGTCACTGG 240
Qy 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
Db 241 TTCACCTCTGGAGCCCTGCTAGACTGTGCCGACATGTGCCGACTTTCTCGGAGGG 300
Qy 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
Db 301 AACCCCAACCTCAGTCTGAGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360
Qy 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
Db 361 GCTGAGCCCGAGGGCTGCGCGGCTGCACCGCGCGGGTGCANAATAGCATCATGACC 420
Qy 141 PheLysAspTyrPheTyrCysTyrAsnThrPheValGluAsnHisGluArgThrPheLys 160
Db 421 TTCAAGATATATTTTACTGCTGGAATACATTTGTAGAAAACCAATGAAAGAACTTCAAA 480
Qy 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGln 175
Db 481 GCCTGGGAAGGCTGCATGAATAATTCAGTTCGTCTCTCCAGACAG 525

RESULT 15
AF529831
LOCUS
DEFINITION Mus musculus clone 4 transgenic Homo sapiens AID (AID) mRNA,
complete cds.
ACCESSION AF529831
VERSION AF529831.1 GI:22297249
KEYWORDS
```

```
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 596)
TITLE Martin,A. and Scharff,M.D.
JOURNAL Somatic hypermutation of the AID transgene in B and non-B cells
2 (bases 1 to 596)
AUTHORS Unpublished
TITLE Martin,A. and Scharff,M.D.
JOURNAL Direct Submission
TITLE Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
JOURNAL Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source
1..596
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="4"
/cell_type="hybridoma Pl-5"
/transgenic
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..>596
/gene="AID"
1..>596
/gene="AID"
/codon_start=1
/product="AID"
/protein_id="AAM95418.1"
/db_xref="GI:22297250"
/translacion="MDSLMMNRKFLYQFNVRWAKGRRETYLCYVVKRRDSATSFSL
DFGYLRNKNCHVELLEFLRYISDWLDPGRCYRVTFWTSRSPCYDCAHVADELGRNP
NLSLRIFATRLYFCEDRKAEPEGLRLHRAGVQIAIMTFKDYFCWMTFVENHERTFK
AWEGLEHNSVRLSRQIRILLPLVEVDLDRDAFRTLGL"
BASE COUNT 127 a 163 c 155 g 151 t
ORIGIN

Alignment Scores:
Pred. No.: 2,24e-176 Length: 596
Score: 175.00 Matches: 175
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 88.38% Indels: 0
DB: 12 Gaps: 0

US-09-966-880A-8 (1-198) x AF529831 (1-596)
Qy 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
Db 1 ATGCACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGCTGG 60
Qy 21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40
Db 61 GCTAAGGTCGGCTGAGACCTACTCTGTCTAGTGTGAAGAGCGCTGACAGTGTCTACA 120
Qy 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
Db 121 TCCTTTTCACTGGACTTTGGTTATCTTCGCANTAAGACGCTGCCAGTGGAAATGCTC 180
Qy 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
Db 181 TTCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCCGCTGCTACCGCGTCACTGG 240
Qy 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
Db 241 TTCACCTCTGGAGCCCTGCTAGACTGTGCCGACATGTGCCGACTTTCTCGGAGGG 300
Qy 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
Db 301 AACCCCAACCTCAGTCTGAGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360
Qy 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
Db 361 GCTGAGCCCGAGGGCTGCGCGGCTGCACCGCGCGGGTGCANAATAGCATCATGACC 420
Qy 141 PheLysAspTyrPheTyrCysTyrAsnThrPheValGluAsnHisGluArgThrPheLys 160
Db 421 TTCAAGATATATTTTACTGCTGGAATACATTTGTAGAAAACCAATGAAAGAACTTCAAA 480
Qy 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGln 175
Db 481 GCCTGGGAAGGCTGCATGAATAATTCAGTTCGTCTCTCCAGACAG 525

RESULT 15
AF529831
LOCUS
DEFINITION Mus musculus clone 4 transgenic Homo sapiens AID (AID) mRNA,
complete cds.
ACCESSION AF529831
VERSION AF529831.1 GI:22297249
KEYWORDS
```

```

Db      361 GCTAGCCGAGGGCTCGCGCGCTGCACCGCGCGGGTGCAAAATAGCCATCATGACC 420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      141 PheLysAspTyrPheTyrCysTyrPheValGluAsnHisGluArgThrPheLys 160
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      421 TTCAAGATTATTTTACTGCTGGATACTTTTGTAGAAACCATGAAAGAACTTTCAAA 480
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGln 175
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      481 GCCTGGGAAGGGCTGCATGAAATTCAGTTCTCTCTCCAGACAG 525
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 16
AF529851
LOCUS   AF529851 597 bp mRNA linear SYN 19-AUG-2002
DEFINITION Cricetulus griseus clone 11 transgenic Homo sapiens AID (AID) mRNA, complete cds.
ACCESSION AF529851
VERSION   AF529851.1 GI:22297289
KEYWORDS  Chinese hamster.
SOURCE    Cricetulus griseus
ORGANISM  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetulus.
REFERENCE 1 (bases 1 to 597)
AUTHORS   Martin,A. and Scharff,M.D.
TITLE      Somatic hypermutation of the AID transgene in B cells and non-B cells
JOURNAL    Unpublished
REFERENCE 2 (bases 1 to 597)
AUTHORS   Martin,A. and Scharff,M.D.
TITLE      Direct Submission
JOURNAL    Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES  Location/Qualifiers
            source
            1. 597
                /organism="Cricetulus griseus"
                /db_xref="taxon:10029"
                /clone="11"
                /cell_line="CHO"
                /transgenic
            2. 597
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
            3. 597
                /gene="AID"
            4. 597
                /gene="AID"
                /codon_start=1
                /product="AID"
                /protein_id="AAM95438.1"
                /db_xref="GI:22297290"
                /translation="MDSLMMNRKFLYQFKNVRWAKGRREYLCYVVKRRDSATSFSL
                DFGYLNRKNGCHVELLRLYISDMDLPGRCYRYVTWTSWSPCYDCARHVAFLRGNP
                NLSRITARTLYFCEDRKAEPEGLRLRHAGVQIAIMTFKDYFCWNTFVENHERTEK
                AWEGHNSVRLSRQLRILLPLYEVDLDRDAFRTWGR"
BASE COUNT 128 a 165 c 157 g 147 t
ORIGIN

Alignment Scores:
Pred. No.: 3.46e-173 Length: 597
Score: 172.00 Matches: 172
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 86.87% Indels: 0
Db: 12 Gaps: 0

US-09-966-880A-8 (1-198) x AF529851 (1-597)

Qy      1 MetAspSerLeuLeuMetAsnArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGCACACCCCTCTGATGACCGGGAAGGAGTTCTTTACCACTTCAAAAATGTCGCTGG 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

```

```

Qy      21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgAspSerAlaThr 40
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GCTAAGGGTCGCGTGAGACCTACTGTGTACGTAGTAGAAGAGCGGTGACAGTGTCTACA 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 TCCTTTTCTACGGGACTTTGGTTATCTTCGCAATAGAACGCGCTGCCAGTGGAAATTTGCTC 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      61 PheLeuArgTyrIleSerAspTyrPheLeuAspProGlyArgCysTyrArgValThrTrp 80
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCCGCTGCTACCCGCTCACCTGG 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 TTCACCTCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGGCCGACTTTCTTCGAGGG 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301 AACCCCAACCTCAGTCTGAGGATCTTACCACCGCGCCTCTACTTCTGTGAGGAGCGCAAG 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      361 GCTGAGCCGAGGGCTGCGCGCGCTGCACCGCGCGGGTGCAAAATAGCCATCATGACC 420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      141 PheLysAspTyrPheTyrCysTyrPheValGluAsnHisGluArgThrPheLys 160
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      421 TTCAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAAGAACTTTCAAA 480
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeu 172
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      481 GCCTGGGAAGGGCTGCATGAAATTCAGTTCTGCTCTC 516
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 17
AF529825
LOCUS   AF529825 596 bp mRNA linear PRI 19-AUG-2002
DEFINITION Homo sapiens clone Ramos 11 AID (AID) mRNA, partial cds.
ACCESSION AF529825
VERSION   AF529825.1 GI:22297237
KEYWORDS  human.
SOURCE    Homo sapiens
ORGANISM  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 596)
AUTHORS   Martin,A. and Scharff,M.D.
TITLE      Somatic hypermutation of the AID transgene in B cells and non-B cells
JOURNAL    Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS   Martin,A. and Scharff,M.D.
TITLE      Direct Submission
JOURNAL    Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES  Location/Qualifiers
            source
            1. 596
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="Ramos 11"
            2. 596
                /gene="AID"
            3. 596
                /gene="AID"
                /note="Integrated into Burkitt's lymphoma cell line Ramos"
                /codon_start=1
                /product="AID"
                /protein_id="AAM95412.1"
                /db_xref="GI:22297238"
                /translation="MDSLMMNRKFLYQFKNVRWAKGRREAYLCYVVKRRDSATSFSL
                DFGYLNRKNGCHVELLRLYISDMDLPGRCYRYVTWTSWSPCYDCARHVAFLRGNP
                NLSRITARTLYFCEDRKAEPEGLRLRHAGVQIAIMTFKDYFCWNTFVENHERTEK
                AWEGHNSVRLSRQLRILLPLYEVDLDRDAFRITGL"
BASE COUNT 126 a 164 c 156 g 150 t
ORIGIN

```



```

ACCESSION AF529833
VERSION AF529833.1 GI:22297253
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 596)
AUTHORS Martin.A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B and non-B cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin.A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
Source
1..596
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="6"
/cell_type="hybridoma P1-5"
/transgenic
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..>596
/gene="AID"
1..>596
/gene="AID"
/codon_start=1
/product="AID"
/protein_id="AAM95420.1"
/db_xref="GI:22297254"
/translation="MDSLMMNRKELYQKKNVRWAKGRRETYVCVVKRRDSATSFSL
DFGYLRNKGCHVELLFURYISDWLDGRCYRVTWFTSWSPCYDCARHVAFLRGNP
NLSLRTFARLYFCEDRAEPGLRLHRAGVOIAMTFKDYFCWNTFVENHRTFK
AWEGUHSVRLSRQLRLLPLYEVDLDRDAFRTGL"
BASE COUNT 127 a 163 c 156 g 150 t
ORIGIN
Alignment Scores:
Pred. No.: 5,32e-170 Length: 596
Score: 169.00 Matches: 197
Percent Similarity: 98.99% Conservative: 0
Best Local Similarity: 98.99% Mismatches: 1
Query Match: 85.35% Indels: 2
DB: 12 Gaps: 0
US-09-966-880A-8 (1-198) x AF529833 (1-596)
QY 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTtr 20
Db 1 ATGGACAGCCTCTTGATGACCGGAGGAGTTCTTTACCAATTCAAAATGTCGCGTG 60
QY 21 AlaLysGlyArgArgGluThrTyrLeu-CystTyrValValLysArgArgAspSerAlath 40
Db 61 GCTAAGGTCGCGCTGAGACCTACGT-GRGCTACGTAGTGAAGCGCGTACAGTCTAC 119
QY 40 rSerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeu 60
Db 120 ATCCCTTTTCACGTGGACTTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGAATGCT 179
QY 60 uPheLeuArgTyrIleSerAspTtrAspLeuAspProGlyArgCysTyrArgValThrTr 80
Db 180 CTTCTCCGCTACACTCGACTGGGACCTAGACCTGGCCGCTGCTACCGGCTCACCTG 239
QY 80 pPheThrSerTtrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgL 100
Db 240 GTTCACCTCCTGGAGCCCTGTACGACTGTGCCGACATGTGGCGGACTTTCTCGGAGG 299
QY 100 yAsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLy 120
|||||

```

---

```

Db 300 GAACCCCAACCTCAGTCTGAGGATCTTCAACCCGCGCCTCTACTTCTGTGAGACCGCAA 359
QY 120 sAlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleWetTh 140
|||||
Db 360 GGCTAGCCCGAGGGCTGCGCGCTGCACCCGCGGGGTGCAAAATAGCCATCATGAC 419
QY 140 rPheLysAspTyrPheTyrCysTtrPAsnThrPheValGluAsnHisGluArgThrPheLy 160
|||||
Db 420 CTTCAAGATATTATTTTACTGCTGGAATACTTTTCTAGAAAAACCATGAAGAATTTTCAA 479
QY 160 sAlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLe 180
|||||
Db 480 ACCCTGGGAAGGCTGCATGAAATTCAGTTCTCTCCACAGACGTTCCGCGCATCCT 539
QY 180 uLeuProLeuTyrGluValAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
|||||
Db 540 TTTGGCCCTGTATGAGTTGATGACTTACGAGACCGCATTTCTGTTGGGACTT 594
RESULT 20
AF529835
LOCUS AF529835 596 bp mRNA linear SYN 19-AUG-2002
DEFINITION Mus musculus clone 8 transgenic Homo sapiens AID (AID) mRNA,
complete cds.
ACCESSION AF529835
VERSION AF529835.1 GI:22297257
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 596)
AUTHORS Martin.A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B and non-B cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin.A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
Source
1..596
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="8"
/cell_type="hybridoma P1-5"
/transgenic
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..>596
/gene="AID"
1..>596
/gene="AID"
/codon_start=1
/product="AID"
/protein_id="AAM95422.1"
/db_xref="GI:22297258"
/translation="MDSLMMNRKELYQKKNVRWAKGRRETYVCVVKRRDSATSFSL
DFGYLRNKGCHVELLFURYISDWLDGRCYRVTWFTSWSPCYDCARHVAFLRGNP
NLSLRTFARLYFCEDRAEPGLRLHRAGVOIAMTFKDYFCWNTFVENHRTFK
AWEGUHSVRLSRQLRLLPLYEVDLDRDAFRTGL"
BASE COUNT 127 a 163 c 156 g 150 t
ORIGIN
Alignment Scores:
Pred. No.: 5,32e-170 Length: 596
Score: 169.00 Matches: 197
Percent Similarity: 98.99% Conservative: 0
Best Local Similarity: 98.99% Mismatches: 1
Query Match: 85.35% Indels: 2
DB: 12 Gaps: 0
US-09-966-880A-8 (1-198) x AF529835 (1-596)

```

```

QY 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTTP 20
Db 1 ATGGACAGCCTCTTATGATGAACCGGAGGAGTTCTTTACCAATTCAAAATGTCGCGTGG 60
QY 21 AlaLysGlyArgArgGluThrTyrLeu-CysTyrValValLysArgArgSerAlaTh 40
Db 61 GCTAAGGGTCGGCGTGAGACCTACGT-GTGCTACGTAGTGAAGAGCGGTGACAGTGCTAC 119
QY 40 rSerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeu 60
Db 120 ATCCCTTTTACGGACTTTGGTTATCTTCGCAATAAGAACGGCTGCCAGTGGAAATGCT 179
QY 60 uPheLeuArgTyrIleSerAspTyrAspLeuAspProGlyArgCysTyrArgValThrTr 80
Db 180 CTTCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCCCTGCTACCGGCTCACCTG 239
QY 80 pPheThrSerTyrSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgG 100
Db 240 GTTCACCTCTCTGGAGCCCTGCTAGGACTGTGCCGACATGTGGCGGACTTTCTCGGAGG 299
QY 100 vAsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgly 120
Db 300 GAACCCCAACCTCAGTCTGAGGATCTTACC CGCGGCTCTACTTCTGTGAGACCGCAA 359
QY 120 saLaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetTh 140
Db 360 GGCTGAGCCCGAGGGCTGCGCGGCTGCACCCGCGCGGGTGCAATAGCCATCATGAC 419
QY 140 rPheLysAspTyrPheTyrCysTyrAsnThrPheValGluAsnHisGluArgThrPheLy 160
Db 420 CTTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAAGAAGTTTCAA 479
QY 160 saLatrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLe 180
Db 480 ACCCTGGGAGGCGTGCAATGAAATTCAGTTCTCTCCAGACAGCTTCGGCGGATCCT 539
QY 180 uLeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
Db 540 TTTTCCCTGTATGAGTGTGATGACTTACGAGACGACATTTCTACTTTGGGACTT 594

RESULT 21
AF529856 547 bp mRNA linear SYN 19-AUG-2002
LOCUS Cricetus griseus clone 16 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION complete cds.
ACCESSION AF529856
VERSION 1 GI:22297299
KEYWORDS Chinese hamster.
SOURCE Cricetus griseus
ORGANISM Cricetus griseus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
Cricetus

REFERENCE
AUTHORS Martin.A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL Unpublished
REFERENCE
AUTHORS Martin.A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
Source Location/Qualifiers
1..547
/organism="Cricetus griseus"
/db_xref="taxon:10029"
/clone="16"
/cell_line="CHO"
/transgenic
1..547
/organism="Homo sapiens"

source
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

/db_xref="taxon:9606"
<1..547
/gene="AID"
<1..547
/gene="AID"
/codon_start=2
/product="AID"
/protein_id="AAM95443.1"
/db_xref="GI:22297300"
/translation="VRWAKGRRETYLCYVVKRRDSATSFSLDFUYLRNKNKGCHVELLF
LRYISDWDLDGRCYRVTFWTSWPCYDCARHVADFLRGNPNLSLURITAFILYFCEDR
KAEPGLRLRHAGVQIAITMFKDYFYCNWTFEVENHRTFKAEGLHENSRLSRLR
RIILPLYEVDDLRDAFTWGR"
BASE COUNT 112 a 153 c 147 g 135 t
ORIGIN
Alignment Scores:
Pred. No.: 1..36e-162 Length: 547
Score: 162.00 Matches: 162
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 81.82% Indels: 0
Db: 12 Gaps: 0
US-09-966-880A-8 (1-198) x AF529856 (1-547)
QY 18 ValArgTTPAlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArg 17
Db 2 GTCCCTGGCTAAGGTCGGCTGAGACCTACCTGCTGCTAGTGAAGAGGCTGAC 61
QY 38 SerAlaThrSerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisVal 57
Db 62 AGTGCTACATCCTTTTCTACCTGGACCTTGTGTTATCTTCCCAATAAGAACCGTGTCCACG 121
QY 58 GluLeuLeuPheLeuArgTyrIleSerAspTyrPheLeuAspProGlyArgCysTyrArg 77
Db 122 GAATTCCTCTTCCTCCGCTACATCTCGGACTCGGACCTGAGACCTGGCGCTTGTCTACG 181
QY 78 ValThrTTPPheThrSerTyrPheCysTyrAspCysAlaArgHisValAlaAspPhe 97
Db 182 GTACCTGGTTCACCTCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGCCCACTTTT 241
QY 98 LeuArgGlyAsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGlu 117
Db 242 CTCGAGGAGAACCCCAACCTCAGTCTGAGGATCTTTCAGCGCGCTCTCTTCTGTGAG 301
QY 118 AspArgLysAlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAla 137
Db 302 GACCCGAGGCTGAGCCCGAGGGCTGCGCGGCTGACCGCGGGGTGCCAATATAGCG 361
QY 138 IleMetThrPheLysAspTyrPheTyrCysTyrPheThrPheValIleAsnHisGluArg 157
Db 362 ATCATGACCTCAAGATTTATTTTACTGCTGTAATCTTTTGTAGAAAAACCATGAAAG 421
QY 158 ThrPheLysAlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArg 177
Db 422 ACTTTCAAAGCCTGGGAGGCTGCATGAAATTCAGTTGCTCTCTCCACAGACGCTCGG 481
QY 178 ArgIle 179
Db 482 CGCATC 487

RESULT 22
AF529822 596 bp mRNA linear PBI 19-AUG-2002
LOCUS Homo sapiens clone Ramos 8 AID (AID) mRNA, partial cds.
DEFINITION Homo sapiens clone Ramos 8 AID (AID) mRNA, partial cds.
ACCESSION AF529822
VERSION 1 GI:22297231
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

```

REFERENCE 1 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B
        cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
        Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
    source      Location/Qualifiers
                1..596
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="Ramos 8"
                1..>596
                /gene="AID"
                1..>596
                /gene="AID"
                /note="Integrated into Burkitt's lymphoma cell line Ramos"
                /codon_start=1
                /product="AID"
                /protein_id="AAM95409.1"
                /db_xref="GI:22297232"
                /translation="MDSLMMNRKFLYQFKNVRWAKGRRETYLCYVVKRRDRATSFSL
                DFGYLRNKNCHVELLFRLYISDMWLDPPGRCYRVTFWTSWSPCYDCARHVADEFLRGNP
                NLSLRIFTARLYFCEDRKAPGELRLRLHRAVGQIAIMTFKDYFCWNTEFVENHERTFK
                AWEGLHNSVRLSRQLRILLPLYEVDLDRADFRTLGL"
BASE COUNT 126 a 165 c 155 g 150 t
ORIGIN
Alignment Scores:
Pred. No.:      1.95e-160      Length:      596
Score:          160.00      Matches:      197
Percent Similarity: 98.99%      Conservative: 0
Best Local Similarity: 98.99%      Mismatches: 1
Query Match:      98.81%      Indels:      2
DB:              9      Gaps:      0

US-09-966-880A-8 (1-198) x AF529822 (1-596)

Qy      1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
Db      1 ATGGACAGCCTCTTGATGAACCGGAGGAGTTCCTTTACCAAAATTCAAAAATGTCGCTGG 60

Qy      21 AlalysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSer-AlaThr 40
Db      61 GCTAAGGGTGGCGTGAGACCTACCTGTGTACGTAGTGAAGAGCGGTGACCG-TGCTAC 119

Qy      40 rSerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLe 60
Db      120 ATCTTTTCACTGGACTTGTGTTATCTTCGCAATAAGAACCGGTGCCACCTGGAATTGCT 179

Qy      60 uPheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTr 80
Db      180 CTTCTCTCCGTATACATCTCGACTGGGACCTAGACCTGGCCGTGCTACCGCGCTACCTG 239

Qy      80 pPheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGl 100
Db      240 GTTCACCTCTCGAGCCCTCTGCTACGACTGTGCCGACATGTGCCGACATTTCTGCCGAGG 299

Qy      100 yAsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLy 120
Db      300 GAACCCCAACCTCAGTGTAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAA 359

Qy      120 sAlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetTh 140
Db      360 GGCTGACCCCGAGGGGCTGGCGGGCTGCACCGCGCGGGGTGCAAAATAGCCATCATGAC 419

Qy      140 rPheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLy 160
Db      420 CTTCAAAGATTATTTTTTACTGCTGGAAATCTTTGTAGAAACCATGAAAGAACTTTTCAA 479

```

```

Qy      160 sAlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLe 180
Db      480 AGCCTGGGAAGGCTGTCATGAAATTCAGTTCTCTCCAGACAGCTTCGGCGCATCTT 539

Qy      180 uLeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
Db      540 TTTGCCCTCTATGAGGTTGATGACTTACGAGAGCGCATTCGTACTTTGGGACTT 594

RESULT 23
AF529834      Mus musculus clone 7 transgenic Homo sapiens AID (AID) mRNA,
LOCUS complete cds.
DEFINITION AF529834
ACCESSION AF529834.1 GI:22297255
VERSION AF529834.1
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 595)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B and non-B cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 595)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
        Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
    source      Location/Qualifiers
                1..595
                /organism="Mus musculus"
                /db_xref="taxon:10090"
                /clone="7"
                /cell_type="hybridoma pl-5"
                /transgenic
                1..595
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                1..>595
                /gene="AID"
                1..>595
                /gene="AID"
                /codon_start=1
                /product="AID"
                /protein_id="AAM95421.1"
                /db_xref="GI:22297256"
                /translation="MDSLMMNRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSL
                DFGYLRNKNCHVELLFRLYISDMWLDPPGRCYRVTFWTSWSPCYDCARHVADEFLRGNP
                NLSLRIFTARLYFCEDRKAPGELRLRLHRAVGQIAIMTFKDYFCWNTEFVENHERTFK
                AWEGLHNSVRLSRQLRILLPLYEVDLDRADFRTLGL"
BASE COUNT 127 a 163 c 154 g 151 t
ORIGIN

Alignment Scores:
Pred. No.:      9.57e-149      Length:      595
Score:          149.00      Matches:      197
Percent Similarity: 98.99%      Conservative: 0
Best Local Similarity: 98.99%      Mismatches: 1
Query Match:      75.25%      Indels:      2
DB:              12      Gaps:      0

US-09-966-880A-8 (1-198) x AF529834 (1-595)

Qy      1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
Db      1 ATGGACAGCCTCTTGATGAACCGGAGGAGTTCCTTTACCAAAATTCAAAAATGTCGCTGG 60

Qy      21 AlalysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSer-AlaThr 40
Db      61 GCTAAGGGTGGCGTGAGACCTACCTGTGTACGTAGTGAAGAGCGGTGACAGTGCATACA 120

Qy      41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60

```

```
Db 121 TCCCTTTTACGGGACTTTGGTTATCTTCGCAATAAGACGGCTGCCACGTGGAAATTCCTC 180
Qy 61 PheLeuArgTyrIleSerAspTyrAspLeuAspProGlyArgCysTyrArgValThrTrp 80
Db 181 TTCTCTCCGCTACATCTCGGACTGGAGCTAGACCTCGCGCTGCTACCGCGTCACTCGG 240
Qy 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
Db 241 TTCACCTCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGCCGCGACTTTCGCGAGG 300
Qy 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
Db 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTCTCTGTGAGGACCGAAG 360
Qy 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
Db 361 GCTGAGCCCGAGGGCTGCGCGCTGCACCGCGCGGGTGCAGAAATAGGCATCATGACC 420
Qy 141 PheLysAspTyrPheTyrCysTyrAsnThr-PheValGluAsnHisGluArgThrPheLys 160
Db 421 TTCAAGATATTATTTTACTGCTGGAATAT-TTTTGTAGAAACCATGAAAGAACTTTCAA 479
Qy 160 SaluTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLe 180
Db 480 AGCCTGGGAGGGCTGCATGAAATTCAGTTCCTCTCCAGACAGCTTCGGCGCATCCT 539
Qy 180 uLeuProLeuTyrGluValAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
Db 540 TTGCCCCCTGTATGAGGTTGATGACTTACGAGACGCAATTCGTACTTTGGGACTT 594

RESULT 24
AF529836 596 bp mRNA linear SYN 19-AUG-2002
LOCUS Mus musculus clone 9 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION complete cds.
ACCESSION AF529836
VERSION AF529836.1 GI:22297259
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
JOURNAL Somatic hypermutation of the AID transgene in B and non-B cells
2 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
JOURNAL Direct Submission
TITLE Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source Location/Qualifiers
1..596
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="9"
/cell_type="hybridoma pl-5"
/transgenic
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..>596
/gene="AID"
1..>596
/gene="AID"
/codon_start=1
/product="AID"
/protein_id="AA095423.1"
/db_xref="GI:22297260"
/translation="MDSLMNRKFLYQFKNVRWSKRRRTYLCYVVKRRDSATPSL
DFGLRNNGCHVLEFLAYISDWLDGRCYRVTFWTSPCYCDARHVAFLRGNP
NLSLRFTARLFCEDRKAEPGLRLRHAGVOIAITWTFNDYFCWNTVEHRTFK
```

```
BASE COUNT 127 a 164 c 153 g 152 t
ORIGIN
AWGLHENSVRLSROLRRILLPLYEVDLDAFRTLGL*
Alignment Scores:
Pred. No.: 1.48e-145 Length: 596
Score: 146.00 Matches: 176
Percent Similarity: 98.88% Conservative: 0
Best Local Similarity: 98.88% Mismatches: 1
Query Match: 73.74% Indels: 2
DB: 12 Gaps: 0

US-09-966-880A-8 (1-198) x AF529836 (1-596)
Qy 22 LysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThrSer 41
Db 64 AAGGTCGGCGTGAGACCTACCTGTCTAGTGAAGAGCGGTGACATGCTACATACC 123
Qy 42 PheSerLeuAspPheGlyTyrLeuArgAsnLys-AsnGlyCysHisValGluLeuLeuPhe 61
Db 124 TTTTCACTGGACTTTGGTTATCTTCGCAATAA-TAACGGCTGCCACGTGCAATTGCTCTT 182
Qy 61 eLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrpPh 81
Db 183 CCTCCGCTACATCTCGGACTGGACCTAGACCTGGCCGCTGCTACCGCTTCACCTGGTT 242
Qy 81 eThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGlyAs 101
Db 243 CACCTCTCGAGGCCCTGCTACGACTGTGCCGACATGTGCCGACTTTTGGCAGGGAA 302
Qy 101 nProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLysAl 121
Db 303 CCCCAACCTCAGTCTCAGGATCTTCACCGCGCGCTCTACTTCTGTGAGACCGCAAGGC 362
Qy 121 aGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThrPh 141
Db 363 TGAGCCCGAGGGCTGCGGGCGGTGCACCGCGCGGGTGCACCAATAGCCATCATGACTT 422
Qy 141 eLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLysAl 161
Db 423 CAAAGATATTATTTTACTGCTGGAATACTTTGTAGAAACCATGAAAGAACTTTCAANGC 482
Qy 161 aTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeuLe 181
Db 483 CTGGGAAGGCTCATGATAAATTCAGTTCCTCTCTCCAGACAGCTTCGGGCGATCCTTTT 542
Qy 181 uProLeuTyrGluValAspAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
Db 543 GCCCCTGTATGAGGTTGATGACTTACGAGACGCAATTTTCGTACTTTGGGACTT 594

RESULT 25
AF529853 597 bp mRNA linear SYN 19-AUG-2002
LOCUS Cricetulus griseus clone 13 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION complete cds.
ACCESSION AF529853
VERSION AF529853.1 GI:22297293
KEYWORDS Chinese hamster.
SOURCE Cricetulus griseus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
Cricetulus.
REFERENCE 1 (bases 1 to 597)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL Unpublished
2 (bases 1 to 597)
AUTHORS Martin,A. and Scharff,M.D.
JOURNAL Direct Submission
TITLE Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
```



Db	540	TTTGCCCTGTATGAGGTTGATGACTTACGAGACGCAATTCGTACT	585
FEATURES	Location/Qualifiers		
source	1..597		
	/organism="Cricetulus griseus"		
	/db_xref="taxon:10029"		
	/clone="13"		
	/cell_line="CHO"		
	/transgenic		
source	1..597		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
gene	1..597		
	/gene="AID"		
CDS	1..597		
	/gene="AID"		
	/codon_start=1		
	/product="AID"		
	/protein_id="AAM95440.1"		
	/db_xref="GI:22297294"		
	/translation="MDSLLMNRKFLYQKNVRWAKGRRETYLCYVVKRRDSATSFSL		
	DFGLRNKNGCHVELLFURYISDMDLDPGRCYRVTFWSPCYDCARHVAADFRLGNP		
	NLSLRIFTARLYFCEDRKAEPGLRLRHAGVQIAIMTFKDYFYWNTFVENHETFK		
	AWEGHENSVRSLRQLRILLPLVEVDLDRDAFTWGR"		
BASE COUNT	129 a	164 c	156 g 148 t
ORIGIN			
Alignment Scores:			
Pred. No.:	1..48e-145	Length:	597
Score:	146.00	Matches:	194
Percent Similarity:	98.98%	Conservative:	0
Best Local Similarity:	98.98%	Mismatches:	1
Query Match:	73.74%	Indels:	2
DB:	12	Gaps:	0
US-09-966-880A-8 (1-198) x AF529853 (1-597)			
QY	1	MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp	20
Db	1	ATGGACAGCCTCTTATGATGAACCGGAGGAAGTTTCTTTACCAATTCAAAAATGTCCGCTGG	60
QY	21	AlaLysGlyArgArgGluThrTyrLeuCystYrValValLysArgArgAspSerAlaThr	40
Db	61	GCTAGGGTCGGCGTGACACCTACTGTCTAGCTAGTAGAAGCGGTGACAGTGTCTACA	120
QY	41	SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeu	60
Db	121	TCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGAACGGCTGCCAGTGGAAATTCCTC	180
QY	61	PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp	80
Db	181	TTCTCCCGCTACATCTCGACTGGGACCTAGACCTGTGGCCGTGCTACCGCGTCACCTGG	240
QY	81	PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly	100
Db	241	TTACCTCTCTGGAGCCCTGTCTACACATGTGCCCGACATGTGGCCGACTTTCTGCGAGGG	300
QY	101	AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys	120
Db	301	AAACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCCTCTACTTCTGTGAGGACCGGAAG	360
QY	121	AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr	140
Db	361	GCTAGCCCGAGGGGCTGCGCGGCTGCACCGCGCGGGGTGCAAAATAGCCATCATGACC	420
QY	141	PheLysAspTyrPheTyrCys-TrpAsnThrPheValGluAsnHisGluArgThrPheLy	160
Db	421	TTCAAGATATATTTTACTA-CTGGAATACTTTTGTAGAAAACCATGAAAGACTTTTCAA	479
QY	160	sAlaTtpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLe	180
Db	480	AGCCTGGGAAGGCTGCATGAAATTCAGTTTCAGTCTCTCCACAGACGCTTGGCGCATCCT	539
QY	180	uLeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThr	195



```

source
1..597
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..597
/feature="AID"
1..597
/feature="AID"
1..597
/codon_start=1
/product="AID"
/protein_id="AAM95428.1"
/db_xref="GI:22297270"
DFGYLRNKGCHVLELLFLRYISDMDLDPGRCYRVWTFWSPCYDCARHVAADFRLGNP
NLSLRITFARLYFCEDRKAEPGLRLHRAGVQIAIMTKDYFCWNTFVENHETFK
AWEGLHNSVRLSQLRILLPLVEVDLDRDFAFTWGR"
BASE COUNT 129 a 164 c 157 g 147 t
ORIGIN

Alignment Scores:
Pred. No.: 3.52e-139 Length: 597
Score: 140.00 Matches: 194
Percent Similarity: 98.98% Conservative: 0
Best Local Similarity: 98.98% Mismatches: 1
Query Match: 70.71% Indels: 2
DB: 12 Gaps: 0

US-09-966-880A-8 (1-198) x AF529841 (1-597)

QY 1 MetAspSerLeuLeuMetAsnArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
DB 1 ATGACAGCCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGCGTGG 60

21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40
DB 61 GCTAAGGTCGCGGTGAGACCTACTGTGCTACGTAGTGAAGAGCGGTGACAGTGTCTACA 120

41 SerPheSerLeuAspPhePheThrAlaArgLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
DB 121 TCCCTTTTCACTGGACTTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGACCGCAAG 360

61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
DB 181 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCGCTGCTACCGGCTCACCTGG 240

81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
DB 241 TTCACCTCTGGAGCCCTGCTACGACTGTGCGCGACATGTGGCGACTTTCTCGGAGGG 300

101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
DB 301 AACCCCAACCTCAGTCTCAGGATCTTCACCGCGCGCTCTACTTCTGTGAGACCGCAAG 360

121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
DB 361 GCTGAGCCGAGGGGCTGCGCGGCTGACCGCGCGGGGTGCAAAATAGCCATCATGACC 420

141 Phe-LysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLy 160
DB 421 AT-CANAGATTATTTTACTGCTGGAATACITTTGTAGAAACCACTGAAAGAACTTTCAA 479

160 sAlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLe 180
DB 480 AGCCTGGGAAGGCTGCATGAAATTCAGTTTCGTCTCTCCAGACAGCTTCGCGCATCCT 539

180 uLeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThr 195
DB 540 TTGCCCCCTGTATGAGGTTGATGACTTACGAGAGCGCATTTCTGCTACT 585

RESULT 29
AF529830 596 bp mRNA linear SYN 19-AUG-2002
LOCUS Mus musculus clone 3 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION

```

```

complete cds.
ACCESSION AF529830
VERSION AF529830.1 GI:22297247
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B and non-B cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submision
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
location/Qualifiers
1..596
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="3"
/cell_type="hybridoma PI-5"
/transgenic
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..596
/feature="AID"
1..596
/feature="AID"
/codon_start=1
/product="AID"
/protein_id="AAM95417.1"
/db_xref="GI:22297248"
/translation="MDSLMMNRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSL
DFGYLRNKGCHVLELLFLRYISDMDLDPGRCYRVWTFWSPCYDCARHVAADFRLGNP
NLSLRITFARLYFCEDRKAEPGLRLHRAGVQIAIMTKDYFCWNTFVENHETFK
AWEGLHNSVRLSQLRILLPLVEVDLDRDFAFTLGL"
BASE COUNT 126 a 164 c 155 g 151 t
ORIGIN

Alignment Scores:
Pred. No.: 4.06e-138 Length: 596
Score: 139.00 Matches: 197
Percent Similarity: 98.99% Conservative: 0
Best Local Similarity: 98.99% Mismatches: 1
Query Match: 70.20% Indels: 2
DB: 12 Gaps: 0

US-09-966-880A-8 (1-198) x AF529830 (1-596)

QY 1 MetAspSerLeuLeuMetAsnArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
DB 1 ATGACAGCCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGCGTGG 60

21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40
DB 61 GCTAAGGTCGCGGTGAGACCTACTGTGCTACGTAGTGAAGAGCGGTGACAGTGTCTACA 120

41 SerPheSerLeuAspPhePheThrAlaArgLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
DB 121 TCCCTTTTCACTGGACTTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGACCGCAAG 360

61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
DB 181 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCGCTGCTACCGGCTCACCTGG 240

81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
DB 241 TTCACCTCTGGAGCCCTGCTACGACTGTGCGCGACATGTGGCGACTTTCTCGGAGGG 300

101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120

```

```

Db      301 AACCCCAACCTCAGTCTGAGGATCTTCAACCGCGGCTCTACTTCTGTGAGACCGCAAG 360
Qy      121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
Db      361 GCTGAGCCGAGGGCTGCGGCGCTGCACCGCGCGGGTGCAATAGCCATCATGTC- 419
Qy      141 -PheLysAspTyrPheTyrCysTyrPheValGlnHisGluArgThrPheLy 160
Db      420 CTTCAAGATATATTTTACTGCTGGAATACTTTGTAGAAAACCAAGAACTTTCAA 479
Qy      160 salATroGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLe 180
Db      480 AGCCTGGGAGGCTGCGATGAAATTCAGTTCTCTCCAGACAGCTTCGGCGCATCCT 539
Qy      180 uLeuProLeuTyrGluValAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
Db      540 TTTGCCCTGTATGAGCTTACGAGACGATTCGTACTTTGGGACTT 594

RESULT 30
AF529838 596 bp mRNA linear SYN 19-AUG-2002
LOCUS Mus musculus clone 11 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION complete cds.
ACCESSION AF529838
VERSION AF529838.1 GI:22297263
KEYWORDS house mouse.
SOURCE Mus musculus.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B and non-B cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source 1..596
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="11"
/cell_type="hybridoma PI-5"
/transgenic
source 1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/gene="AID"
1..596
/gene="AID"
1..596
/codon_start=1
/product="AID"
/protein_id="AAM95425.1"
/db_xref="GI:22297264"
/translaton="MDSLLMNRKFLYQFNVRWAKGRRETYLCYVVKRRDSATSFSL
DFGLRNKNGCHEVLEFLRYISDWLDPPGRYVTFWFTSWSPCYDCARHVDFLRGNP
NLSIRIFARLYCECDKAEPEGLRLHRRAGVOIVMTFKDYFCWNTFFVENHERTFK
AWELHENSVRLSRLRILLPLYEVDDLRDARTLGL"
BASE COUNT 128 a 163 c 154 g 151 t
ORIGIN

Alignment Scores:
Pred. No.: 6.27e-135 Length: 596
Score: 136.00 Matches: 197
Percent Similarity: 98.99% Conservative: 0
Best Local Similarity: 98.99% Mismatches: 1
Query Match: 68.69% Indels: 2
DB: 12 Gaps: 0

```

```

US-09-966-880A-8 (1-198) x AF529838 (1-596)
Qy      1 MetAspSerLeuLeuMetAsnArgLysPheLeuTyrGlnPheLysAsnValArgTrrp 20
Db      1 ATGACAGCCTCTTGTATGCAACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGGCTGG 60
Qy      21 AlalysGlyArgArgGluThrTyrLeuCysTyrValValLysArgAspSerAlaThr 40
Db      61 GCTAAGGTCGCGCTGAGACCTACCTCTGCTACCTAGTGAAGAGCGGTGACAGTGTTACA 120
Qy      41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
Db      121 TCCCTTTTCACTGGAGCTTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGAATTCCTC 180
Qy      61 PheLeuArgTyrIleSerAspTrrpAspLeuAspProGlyArgCysTyrArgValThrTrrp 80
Db      181 TTTCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGCCCGCTGCTACCTACCTCCTGG 240
Qy      81 PheThrSerTrrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
Db      241 TTCACCTCTCTGGAGCCCTGCTAGCACTGTGCCGACATGTGCCGCGACTTTCTGCGAGGG 300
Qy      101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
Db      301 AACCCCAACCTCAGTCTGAGGATCTTCAACCGCGGCTCTACTTCTGTGAGACCGCAAG 360
Qy      121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
Db      361 GCTGAGCCGAGGGCTGCGGCGCTGCACCGCGCGGGTGCAATAGT-CATCATGAC 419
Qy      140 rPheLysAspTyrPheTyrCysTrrpAsnThrPheValGlnHisGluArgThrPheLy 160
Db      420 CTTCAAGATATATTTTACTGCTGGAATACTTTGTAGAAAACCAAGAACTTTCAA 479
Qy      160 salATrrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLe 180
Db      480 AGCCTGGGAGGCTGCGATGAAATTCAGTTCTCTCCAGACAGCTTTGGCGCACTT 539
Qy      180 uLeuProLeuTyrGluValAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
Db      540 TTTGCCCTGTATGAGCTTACGAGACGATTCGTACTTTGGGACTT 594

RESULT 31
AF529845 597 bp mRNA linear SYN 19-AUG-2002
LOCUS Cricetulus griseus clone 5 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION complete cds.
ACCESSION AF529845
VERSION AF529845.1 GI:22297277
KEYWORDS Chinese hamster.
SOURCE Cricetulus griseus
ORGANISM Cricetulus griseus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
Cricetulus.
REFERENCE 1 (bases 1 to 597)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 597)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source 1..597
/organism="Cricetulus griseus"
/db_xref="taxon:10029"
/clone="5"
/cell_line="CHO"
/transgenic
source 1..597

```



```

QY      140 rPhelYsAspTyrPheTyrCysTyrAsnThrPheValGluAsnHisGluArgThrPhely 160
      |||
      420 CTTCAAGATATATTTTACTGCTGGAATACCTTTGTAGAAAACCATGAAGAACTTTCAA 479
QY      160 sAlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgileLe 180
      |||
      480 AGCCTGGGAAGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTCGGCGCATCCT 539
QY      180 uLeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
      |||
      540 TTTGCCCTGTATGAGTGTAGTACTTACGAGACCATTTCTGTTGGGACTT 594

RESULT 33
AF529848
LOCUS      AF529848      597 bp      mRNA      linear      SYN 19-AUG-2002
DEFINITION Cricetulus griseus clone 8 transgenic Homo sapiens AID (AID) mRNA,
            complete cds.
ACCESSION AF529848
VERSION   AF529848.1 GI:22297283
KEYWORDS  .
SOURCE    Chinese hamster.
ORGANISM  Cricetulus griseus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
            Cricetulus.
REFERENCE 1 (bases 1 to 597)
AUTHORS   Martin,A. and Scharff,M.D.
TITLE     Somatic hypermutation of the AID transgene in B cells and non-B
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 597)
AUTHORS   Martin,A. and Scharff,M.D.
TITLE     Direct Submission
JOURNAL   Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
            Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
            source
            1..597
            /organism="Cricetulus griseus"
            /db_xref="taxon:10029"
            /clone="8"
            /cell_line="CHO"
            /transgenic
            1..597
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            1..597
            /gene="AID"
            1..597
            /gene="AID"
            /codon_start=1
            /product="AID"
            /protein_id="AA095435.1"
            /db_xref="GI:22297284"
            /translation="MDSLMMRRFLYQFKNVRWAKGRRETYLCYVKKRRDSATFSFL
            DFGYLRNKGCHVELLYRISDMGLDPGRYRVYFTWTSWSPCYDCARHVAADFRLRNP
            NLSRIFARLYFEDRAEPEGLRLRHAGVOIAIMTFKDYFCWNTFVENHETFK
            AWGLHENSVRSLRRLPLVEVDLDRDAFTWGR"
BASE COUNT 127 a 164 c 158 g 148 t
ORIGIN
Alignment Scores:
Pred. No.:      2,66e-124      Length:      597
Score:          126.00      Matches:      194
Percent Similarity: 98.98%      Conservative: 0
Best Local Similarity: 98.98%      Mismatches: 1
Query Match:      63.64%      Indels: 2
DB:              12      Gaps: 0

US-09-966-880A-8 (1-198) x AF529848 (1-597)

QY      1 MetAspSerLeuLeuMetAsnArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
      |||

```

```

Db      1 ATGCAGAGCCTCTTGATGAACCGGAGGAAGTTTCTTTACCAATTCAAAAATCGCGCTGG 60
QY      21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgAspSerAlaThr 40
      |||
      61 GCTAAGGTCGGCGTGAGACCTTACCTGCTACGTACGTACGTAGTGAAGAGCGGTGACAGTCTACA 120
QY      41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
      |||
      121 TCCTTTTTCACGTGAGCTTTGGTTATCTTCGCAATAAGAACGGCTGCCACCTGGAATTCGTC 180
QY      61 PheLeuArgTyrIleSerAspTrpAsp-LeuAspProGlyArgCysTyrArgValThrTr 80
      |||
      181 TTCTCCCGCTACATCTCGACTGGGG-CCTAGACCCCTGGCCGCTGCTACCGGCTCACCTG 239
QY      80 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgQ 100
      |||
      240 GTTCACCTCCCTGGAGCCCTGCTACGACTGTGCCGACATGTGGCGGACTTTCTGCGAGG 299
Db      100 YAsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArg 120
      |||
      300 GAACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGACCGCAA 359
QY      120 sAlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetTh 140
      |||
      360 GGCTGAGCCCGAGGGGCTCGCGGCTGCACCGCGCGGGTGCANAATAGTCATCATGAC 419
Db      140 rPhelYsAspTyrPheTyrCysTyrAsnThrPheValGluAsnHisGluArgThrPhely 160
      |||
      420 CTTCAAGATATATTTTACTGCTGGAATACCTTTGTAGAAAACCATGAAGAACTTTCAA 479
QY      160 sAlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgileLe 180
      |||
      480 AGCCTGGGAAGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTCGGCGCATCCT 539
QY      180 uLeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThr 195
      |||
      540 TTTGCCCTGTATGAGTGTAGTACTTACGAGACGCATTTTCGTACT 585

RESULT 34
AF529839
LOCUS      AF529839      596 bp      mRNA      linear      SYN 19-AUG-2002
DEFINITION Mus musculus clone 12 transgenic Homo sapiens AID (AID) mRNA,
            complete cds.
ACCESSION AF529839
VERSION   AF529839.1 GI:22297265
KEYWORDS  .
SOURCE    house mouse.
ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            Martin,A. and Scharff,M.D.
            Somatic hypermutation of the AID transgene in B and non-B cells
            Unpublished
JOURNAL   Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
            Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
            Location/Qualifiers
            1..596
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /clone="12"
            /cell_type="hybridoma p1-5"
            /transgenic
            1..596
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            1..596
            /gene="AID"
            1..596
            /gene="AID"

```

```
/codon_start=1
/product="AID"
/protein_id="AAM95426.1"
/db_xref="GI:22297266"
/translation="MDSLMMRRKFLYQFNKVRKGRRETYLQYVVKRRDSATSFSL
DFGYLRKNKGCHVELLFRLYISDWDLDGRCYRVYTWTSNSPCYDCARHVAFLRGNP
NLSLRIFTARLYFCELDKRAEPLRLHRAVGQVIAIMTFKDYFCWNTFVENHETERK
AMEGLHNSVRLSRQLRLLPLYEVDLDRDAFTLGL"
BASE COUNT      128 a  163 c  155 g  150 t
ORIGIN
Alignment Scores:
Pred. No.:      4.74e-120      Length:      596
Score:          122.00      Matches:      197
Percent Similarity: 98.98%      Conservative: 0
Best Local Similarity: 98.99%      Mismatches: 1
Query Match:      61.62%      Indels:      2
DB:              12      Gaps:      0

US-09-966-880A-8 (1-198) x AF529839 (1-596)
Qy      1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
Db      1 ATGGACACCCCTCTTGATGAACCGGAGGAGTTCTTTACCAATTCAAAAATGTCGCGTGG 60
Qy      21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40
Db      61 GCTAAGGGTCGCGGTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGCTACA 120
Qy      41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
Db      121 TCCCTTTTTCACGTGGACTTTGGTTATCTTCGCCAATAAGAACGCGCTGCCACGTGGAAATGGTC 180
Qy      61 PheLeuArgTyrIleSerAspTyrAspLeuAspProGlyArgCysTyrArgValThrTrp 80
Db      181 TTCCTCCGCTACATCTCGGACTGGGACCTACACCTGGCCGCTGCTACCGGCTCACCTGG 240
Qy      81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
Db      241 TTCACCTCCTGGAGCCCTGCTACGACGTGTCGCCGACATGTGGCGGACTTTCTGCGAGGG 300
Qy      101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
Db      301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGACCGCAAG 360
Qy      121 AlaGluPro-GluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetTh 140
Db      361 CTGAGAC-CAGGGGGCTGCGGGCTGCACCGCGCGGGTGCNAATAGCCATCATGAC 419
Qy      140 rPheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLy 160
Db      420 CTTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAAGAACTTTTCAA 479
Qy      160 salatrPgluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLe 180
Db      480 AGCCTGGGAAGGGTGCATGAAATTCAGTTTCGTCTCTCCAGACAGCTTCGCGCATCCT 539
Qy      180 uLeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
Db      540 TTTGCCCTGTATGAGTTGATGACTTACGAGACGCAATTCGTACTTTGGGACATT 594

RESULT 35
AF529850
LOCUS
DEFINITION
Cricetulus griseus clone 10 transgenic Homo sapiens AID (AID) mRNA,
complete cds.
ACCESSION
AF529850
VERSION
AF529850.1
KEYWORDS
GI:22297287
SOURCE
Chinese hamster.
ORGANISM
Cricetulus griseus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
```

```
Cricetulus.
1 (bases 1 to 597)
Martin,A. and Scharff,M.D.
Somatic hypermutation of the AID transgene in B cells and non-B
cells
Unpublished
REFERENCE
2 (bases 1 to 597)
Martin,A. and Scharff,M.D.
Direct Submission
JOURNAL
Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
Location/Qualifiers
1..597
/organism="Cricetulus griseus"
/db_xref="taxon:10029"
/clone="10"
/cell_line="CHO"
/transgenic
1..597
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..597
/gene="AID"
1..597
/gene="AID"
/codon_start=1
/product="AID"
/protein_id="AAM95437.1"
/db_xref="GI:22297288"
/translation="MDSLMMRRKFLYQFNKVRKGRRETYLQYVVKRRDSATSFSL
DFGYLRKNKGCHVELLFRLYISDWDLDGRCYRVYTWTSNSPCYDCARHVAFLRGNP
NLSLRIFTARLYFCELDKRAEPLRLHRAVGQVIAIMTFKDYFCWNTFVENHETERK
AMEGLHNSVRLSRQLRLLPLYEVDLDRDAFTWGR"
BASE COUNT      128 a  164 c  156 g  149 t
ORIGIN
Alignment Scores:
Pred. No.:      4.75e-120      Length:      597
Score:          122.00      Matches:      194
Percent Similarity: 98.98%      Conservative: 0
Best Local Similarity: 98.98%      Mismatches: 1
Query Match:      61.62%      Indels:      2
DB:              12      Gaps:      0

US-09-966-880A-8 (1-198) x AF529850 (1-597)
Qy      1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
Db      1 ATGGACACCCCTCTTGATGAACCGGAGGAGTTCTTTACCAATTCAAAAATGTCGCGTGG 60
Qy      21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40
Db      61 GCTAAGGGTCGCGGTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGCTACA 120
Qy      41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
Db      121 TCCCTTTTTCACGTGGACTTTGGTTATCTTCGCCAATAAGAACGCGCTGCCACGTGGAAATGGTC 180
Qy      61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGly-ArgCysTyrArgValThrTr 80
Db      181 TTCCTCCGCTACATCTCGGACTGGGACCTACGACCTGT-CCGCTGCTACCGCGTCACCTG 239
Qy      80 pPheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgG1 100
Db      240 GTTACCTCTCTGGAGCCCTCTGCTACGACTGTGCCGACATGTGGCCGACTTTCTTCGCGAGG 299
Qy      100 yAsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLy 120
Db      300 GAACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGCTGAGGACCGCAA 359
Qy      120 salAGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetTh 140
Db      360 GGCTGAGCCCGGAGGGCTGCGCGCGCTGCACCGCGCGGGTGCATAATAGCATCATGAC 419
```

```

QY      140 rPhelysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLy 160
      |||||||
Db      420 CTTCAAGATTATTTTACTGCTGGAATACCTTTGTAGAAACCATGAAGAACTTTCAA 479

QY      160 salaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgileLe 180
      |||||||
Db      480 AGCCTGGGAAGGCTGCGATGAAATTCAGTTGCTCTCCAGACAGCTTCGGCGCATCTCT 539

QY      180 uLeuProLeuTyrGluValAspLeuArgAspAlaPheArgThr 195
      |||||||
Db      540 TTTGCCCTGTATGAGTTGATGACTTACGAGACGCAATTCGTACT 585

RESULT 36
AF529818
LOCUS      Homo sapiens clone Ramos 4 truncated AID (AID) mRNA, complete cds.
DEFINITION
ACCESSION AF529818
VERSION    AF529818.1 GI:22297223
KEYWORDS
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS    Martin,A. and Scharff,M.D.
TITLE      Somatic hypermutation of the AID transgene in B cells and non-B
            cells
JOURNAL    Unpublished
REFERENCE   2 (bases 1 to 596)
AUTHORS    Martin,A. and Scharff,M.D.
TITLE      Direct Submission
JOURNAL    Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
            Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES   source
            1..596
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="Ramos 4"
            /gene="AID"
            /gene="AID"
            /note="Integrated into Burkitt's lymphoma cell line Ramos"
            /codon_start=1
            /product="truncated AID"
            /protein_id="AA095405.1"
            /db_xref="GI:22297224"
            /translation="MDSLMMRRKFLYQFNVRWAKGRRETYLCYVVKRRDSATSFSL
            DFGYLRNKGCHVELLFRLYISDWLDGRCYRVYTWFTSWSPCYDCARHVADFLRGNP
            NLSRIETARLYFCEDHKAEPGLRLHRAQVIAIMTFKDYFCWNTFVENHERTEK
            AWEGLHNSVRLSQLRLLPLVEVDL"
BASE COUNT 128 a 163 c 154 g 151 t
ORIGIN

Alignment Scores:
Pred. No.:      8.45e-116      Length:      596
Score:          118.00      Matches:      188
Percent Similarity: 98.95%      Conservative: 0
Best Local Similarity: 98.95%      Mismatches: 1
Query Match:      59.60%      Indels:      2
DB:              9      Gaps:      0

US-09-966-880A-8 (1-198) x AF529818 (1-596)

QY      1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
      |||||||
Db      1 ATGGACAGCCCTTGTATGAACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGGCTGG 60

QY      21 AlaLysGlyArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40
      |||||||
Db      61 GCTAAGGGTCGGCTGAGACCTTACTGTGCTACGTAGTAGAAGAGCGGTGACAGTGCTACA 120

```

```

QY      41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
      |||||||
Db      121 TCCTTTTCTACTGGACTTTGGTTATCTTCCCAATAAGAACGGCTGCCACCTGGAAATGCTC 180

QY      61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
      |||||||
Db      181 TTCCCTCCGCTACATCTCGACTGGGACCTAGACCCTGGCCGCTGCTACGCGTCACTGG 240

QY      81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
      |||||||
Db      241 TTCACCTCTCGAGCCCTGCTACGACTGTGCCCGACATGTGCCGCACTTTCTGCGAGGG 300

QY      101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArg-Ly 120
      |||||||
Db      301 AACCCCAACCTCAGTCTCAGGATCTTACCCGCGCCCTCTACTTCTGTGAGAGCCA-CAA 359

QY      120 salaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetTh 140
      |||||||
Db      360 GGCTGAGCCCGAGGGGCTGCGCGGCTGCACCGCGCGGTGCAATAGCCATCATGAC 419

QY      140 rPhelysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLy 160
      |||||||
Db      420 CTTCAAGATTATTTTACTGCTGGAATACCTTTCTAGAAACCATGAAGAACTTTCAA 479

QY      160 salaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgileLe 180
      |||||||
Db      480 AGCCTGGGAAGGCTGCGATGAAATTCAGTTGCTCTCCAGACAGCTTCGGCGCATCTCT 539

QY      180 uLeuProLeuTyrGluValAspLeu 189
      |||||||
Db      540 TTTGCCCTGTATGAGTTGATGACTTGA 567

RESULT 37
AF529819
LOCUS      Homo sapiens clone Ramos 5 AID (AID) mRNA, partial cds.
DEFINITION
ACCESSION AF529819
VERSION    AF529819.1 GI:22297225
KEYWORDS
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 596)
AUTHORS    Martin,A. and Scharff,M.D.
TITLE      Somatic hypermutation of the AID transgene in B cells and non-B
            cells
JOURNAL    Unpublished
REFERENCE   2 (bases 1 to 596)
AUTHORS    Martin,A. and Scharff,M.D.
TITLE      Direct Submission
JOURNAL    Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
            Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES   source
            1..596
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="Ramos 5"
            /gene="AID"
            /gene="AID"
            /note="Integrated into Burkitt's lymphoma cell line Ramos"
            /codon_start=1
            /product="AID"
            /protein_id="AA095406.1"
            /db_xref="GI:22297226"
            /translation="MDSLMMRRKFLYQFNVRWAKGRRETYLCYVVKRRDSATSFSL
            DFGYLRNKGCHVELLFRLYISDWLDGRCYRVYTWFTSWSPCYDCARHVADFLRGNP
            NLSRIETARLYFCEDHKAEPGLRLHRAQVIAIMTFKDYFCWNTFVENHERTEK
            AWEGLHNSVRLSQLRLLPLVEVDL"
BASE COUNT 128 a 164 c 154 g 150 t
ORIGIN

```



Alignment Scores:		8.45e-116	Length:	596
Pred. No.:		118.00	Matches:	197
Score:		98.99%	Conservative:	0
Percent Similarity:		98.99%	Mismatches:	1
Best Local Similarity:		59.60%	Indels:	2
Query Match:		9	Gaps:	0
DB:				
US-09-966-880A-8 (1-198) x AF529819 (1-596)				
Qy	1	MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp	20	
Db	1	ATGGACAGCCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGCTGG	60	
Qy	21	AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr	40	
Db	61	GCTAAGGGTCGGCTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGCTACA	120	
Qy	41	SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu	60	
Db	121	TCCTTTTACCTGGAGCTTTCGCAATTAAGAAGCGCTGCCAGCTGGAATTCCTC	180	
Qy	61	PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp	80	
Db	181	TTCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCCGCTGCTACCGCTCACCTGG	240	
Qy	81	PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly	100	
Db	241	TTACCTCTCGAGCCCTGCTAGACTGTGCCGACATGTGGCCGACTTTCTCGGAGGG	300	
Qy	101	AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGlu-AspArgLys	120	
Db	301	AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGT-GGACCGCAA	359	
Qy	120	sAlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetTh	140	
Db	360	GGCTGAGCCCGAGGGGCTGCGCGGCTGCACCGCGCGGTGCAAAATAGGCATCATGAC	419	
Qy	140	rPheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys	160	
Db	420	CTTCAGAGATTATTTTACTGCTGGAATACCTTTTGTAGAAAAACCATGAAAGACTTC	479	
Qy	160	sAlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLe	180	
Db	480	AGCCTGGGAAGGGTGCATGAAATTCAGTTTCGTCTCTCCAGACAGCTTCGCGCATCCT	539	
Qy	180	uLeuProLeuTyrGluValAspLeuArgAspAlaPheArgThrLeuGlyLeu	198	
Db	540	TTTGCCCTGTATGAGGTGATGACTTACGAGACGCAATTCGTACTTTGGGACTT	594	
RESULT 39				
AF529849				
LOCUS				
AF529849				
597 bp				
mRNA				
linear				
SYN 19-AUG-2002				

source		1. .596	/organism="Homo sapiens"	
			/db_xref="taxon:9606"	
			/clone="Ramos 6"	
gene		1. >596		
			/gene="AID"	
CDS		1. >596		
			/gene="AID"	
			/note="Integrated into Burkitt's lymphoma cell line Ramos"	
			/codon_start=1	
			/product="AID"	
			/protein_id="AAM95407.1"	
			/db_xref="GI:22297228"	
			/translation="MDSLIMNRKELYOFKNVRWAKGRRETYLCYVVKRRDSATSFSL	
			DFGYLRNKNGCHVELLLFYRISDMDLPGRCYRVTFWTSMSPCYDCAHRVADFLRGNP	
			NLSRIETARLYFCVDRKAPGPEGLRLHRAGVQIAIMTFKDYFCWNTFEVNHETFK	
			AWELGHSNVRLSRQLRRILLPLYEVDLDRDAFTLGL"	
BASE COUNT	126 a	164 c	155 g	151 t
ORIGIN				
Alignment Scores:				
Pred. No.:	1.13e-113	Length:	596	
Score:	116.00	Matches:	197	
Percent Similarity:	98.99%	Conservative:	0	
Best Local Similarity:	98.99%	Mismatches:	1	
Query Match:	58.59%	Indels:	2	
DB:	9	Gaps:	0	
US-09-966-880A-8 (1-198) x AF529820 (1-596)				
Qy	1	MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp	20	
Db	1	ATGGACAGCCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGCTGG	60	
Qy	21	AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr	40	
Db	61	GCTAAGGGTCGGCTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGCTACA	120	
Qy	41	SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu	60	
Db	121	TCCTTTTACCTGGAGCTTTCGCAATTAAGAAGCGCTGCCAGCTGGAATTCCTC	180	
Qy	61	PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp	80	
Db	181	TTCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCCGCTGCTACCGCTCACCTGG	240	
Qy	81	PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly	100	
Db	241	TTACCTCTCGAGCCCTGCTAGACTGTGCCGACATGTGGCCGACTTTCTCGGAGGG	300	
Qy	101	AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGlu-AspArgLys	120	
Db	301	AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGT-GGACCGCAA	359	
Qy	120	sAlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetTh	140	
Db	360	GGCTGAGCCCGAGGGGCTGCGCGGCTGCACCGCGCGGTGCAAAATAGGCATCATGAC	419	
Qy	140	rPheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys	160	
Db	420	CTTCAGAGATTATTTTACTGCTGGAATACCTTTTGTAGAAACCATGAAAGACTTC	479	
Qy	160	sAlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLe	180	
Db	480	AGCCTGGGAAGGGTGCATGAAATTCAGTTTCGTCTCTCCAGACAGCTTCGCGCATCCT	539	
Qy	180	uLeuProLeuTyrGluValAspLeuArgAspAlaPheArgThrLeuGlyLeu	198	
Db	540	TTTGCCCTGTATGAGGTGATGACTTACGAGACGCAATTCGTACTTTGGGACTT	594	
RESULT 38				
AF529820				
LOCUS				
AF529820				
596 bp				
mRNA				
linear				
PRI 19-AUG-2002				
Homo sapiens				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
1 (bases 1 to 596)				
Martin,A. and Scharff,M.D.				
Somatic hypermutation of the AID transgene in B cells and non-B				
cells				
Unpublished				
2 (bases 1 to 596)				
Martin,A. and Scharff,M.D.				
Direct Submission				
Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of				
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA				
Location/Qualifiers				
FEATURES				

```

DEFINITION Cricetulus griseus clone 9 transgenic Homo sapiens AID (AID) mRNA,
complete cds.
ACCESSION AF529849
VERSION AF529849.1 GI:22297285
KEYWORDS Chinese hamster.
ORGANISM Cricetulus griseus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
Cricetulus.
REFERENCE 1 (bases 1 to 597)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 597)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
    source
        1..597
            /organism="Cricetulus griseus"
            /db_xref="taxon:10029"
            /clone="9"
            /cell_line="CHO"
            /transgenic
        1..597
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            1..597
            /gene="AID"
            1..597
            /gene="AID"
            /product="AID"
            /protein_id="AA095436.1"
            /db_xref="GI:22297285"
            /translation="MDSLLMNRKFLYQFNVRWAKGRRETYLCYVVKRRDSATSFSL
            DFGVLRKNGCHVELFLRYISDMDLPGRCYRYTWTSPCYDCARHVAFLRGNP
            NLSLRITARIYFCGDRKAPEGLRLRHAGVQIAITMFDYFCWNTFVENHERTFK
            AMEGLHNSRLSRLRILLPLVEVDLDRDAFTWGR"
BASE COUNT 127 a 164 c 158 g 148 t
ORIGIN

Alignment Scores:
Pred. No.: 1.13e-113 Length: 597
Score: 116.00 Matches: 194
Percent Similarity: 98.98% Conservative: 0
Best Local Similarity: 98.98% Mismatches: 1
Query Match: 58.59% Indels: 2
DB: 12 Gaps: 0

US-09-966-880A-8 (1-198) x AF529849 (1-597)

Qy 1 MetAspSerLeuLeuMetAsnArgLysPheLeuTyrGlnPheLysAsnValArgTrrp 20
|
Db 1 ATGGACAGCCTCTTGATGAACCGGAGGAGTTCTTTACCAATTCAAAAATGTCGCTGG 60
|
Qy 21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40
|
Db 61 GCTAAGGTCGCGCTGAGACCTACCTGCTGTCTAGTGTAGAGAGCGCTGACAGTGCTACA 120
|
Qy 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
|
Db 121 TCCTTTTCACGGACTTGGTTATCTTCGCAATAAGAACGCTGCCACGTGGAATGCTC 180
|
Qy 61 PheLeuArgTyrIleSerAspTrrpAspLeuAspProGlyArgCysTyrArgValThrTrrp 80
|
Db 181 TTCCTCGCTACATCTCGGACTGGACCTAGACCTGGCCCTGCTACCGGCTCACCTGG 240
|
Qy 81 PheThrSerTrrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
|

```

```

Db 241 TTCACCTCCTCGAGCCCTGCTACGACTGTGCCGACATGTGGCGGACTTTTCTCGCAGG 300
|
Qy 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGlu-AspArgLys 120
|
Db 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCCCTCTACTTCTGTGTG-GGACCGCA 359
|
Qy 120 sAlaGluProGluGlyLeuArgLeuHisArgAlaGlyValGlnIleAlaIleMetTh 140
|
Db 360 GGCTGAGCCCGAGGGGCTGCCGCGCTGCACCGCGGGGTGCAATAATAGCATCATGAC 419
|
Qy 140 rPheLysAspTyrPheTyrCysTrrpAsnThrPheValGluAsnHisGluArgThrPheLys 160
|
Db 420 CTTCAAGATTATTTTACTGCTGGAATACTTTGTAGAAAAACCATGAACAACCTTTTCAA 479
|
Qy 160 sAlaTrrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
|
Db 480 AGCCTGGGAGGGCTGCATGAAAATTCAGTTCTGCTCTCCAGACAGCTTTGGCGCATCCT 539
|
Qy 180 uLeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThr 195
|
Db 540 TTTGCCCTGTATGAGGTGATGACTTACGAGACGCATTTCTGCTACT 585
|

RESULT 40
AF529840 577 bp mRNA linear SYN 19-AUG-2002
LOCUS Mus musculus clone 13 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION complete cds.
ACCESSION AF529840
VERSION AF529840.1 GI:22297267
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
REFERENCE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
AUTHORS Martin,A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B and non-B cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 577)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
    Location/Qualifiers
        1..577
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /clone="13"
            /cell_type="hybridoma PI-5"
            /transgenic
        1..577
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            <1..>577
            /gene="AID"
            <1..>577
            /gene="AID"
            /codon_start=3
            /product="AID"
            /protein_id="AA095427.1"
            /db_xref="GI:22297268"
            /translation="RRKFLYQFNVRWAKGRRETYLCYVVKRRDSATSFSLDFCYLNRN
            KNCHVELLEFLRYISDMDLPGRCYRYTWTSPCYDCARHVAFLRGPNLSLRIF
            TARYLFCEDRKAESGRLRHAGVQIAITMFDYFCWNTFVENHERTFKAMEGLLHE
            NSVRLSRLRILLPLVEVDLDRDAFTLGL"
BASE COUNT 122 a 159 c 150 g 146 t
ORIGIN

Alignment Scores:
Pred. No.: 1.27e-112 Length: 577
Score: 115.00 Matches: 190
Percent Similarity: 98.96% Conservative: 0
Best Local Similarity: 98.96% Mismatches: 1

```

Query Match:	58.08%	Indels:	2
DB:	12	Gaps:	0
US-09-966-880A-8 (1-198) x AF529840 (1-577)			
QY	8	ArgArgLysPheLeuTyrGlnPheLysAsnValArgTrrpAlaLysGlyArgArgGluThr	27
DB	3	CGAGGAAGTTCCTTTACCAATTCAAAAATGCCGTGGCTPAAGGGTCGGCGTGAGACC	62
QY	28	TyrLeuCysTyrValValLysArgArgAspSerAlaThrSerPheSerLeuAspPheGly	47
DB	63	TACCTGTCTAGTAGTAGAGAGCGTGACAGTGTATCATCTTTTCACTGGACTTTGGT	122
QY	48	TyrLeuArgAsnLysAsnGlyCysHisValGluLeuPheLeuArgTyrIleSerAsp	67
DB	123	TATCTTCGCAATAAGAACGGCTGCCACGTGGAATTCCTTCTCCGCTACATCTCGGAC	182
QY	68	TrpAspLeuAspProGlyArgCysTyrArgValThrTrpPheThrSerTrpSerProCys	87
DB	183	TGGGACCTAGACCCCTGGCGCGCTGCTACCGCGTCACCTGGTTCCACTCTGGAGCCCTGC	242
QY	88	TyrAspCysAlaArgHisValAlaAspPheLeuArgGlyAsnProAsnLeuSerLeuArg	107
DB	243	TACGACTGTGCCGACATGTGCCGACTTCTTCGAGGGAACCCCAACCTCAGTCTGAGG	302
QY	108	IlePheThrAlaArgLeuTyrPheCysGluAspArgLysAlaGluPro-GluGlyLeuAr	127
DB	303	ATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAGGCTGAGTC-CGAGGGGCTGCG	361
QY	127	gArgLeuHisArgAlaGlyValGlnIleAlaIleMetThrPheLysAspTyrPheTyrCy	147
DB	362	CGGGCTGCACCGCGCGGGTGCAATAGCCATCATGACCTTCAAGAGATTATTTTACTG	421
QY	147	stirAsnThrPheValGluAsnHisGluArgThrPheLysAlaTrpGluGlyLeuHisGl	167
DB	422	CTGGATACTTTTGTAGAAACCATGAAGAATTTCAAGGCTGGGAGGCTGCATGA	481
QY	167	uAsnSerValArgLeuSerArgGlnLeuArgArgIleLeuLeuProLeuTyrGluValas	187
DB	482	AAATTTCAGTTCCTCTCCAGACAGCTTCGGCGCATCTTTTGCCCTGTATGAGTTGA	541
QY	187	pAspLeuArgAspAlaPheArgThrLeuGlyLeu	198
DB	542	TGACTTACGAGACGCATTTCGTACTTTGGGACTT	575
RESULT 41			
AF529816			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source			
gene			

CDS	1..>596	/gene="AID"	
		/note="integrated into Burkitt's lymphoma cell line Ramos"	
		/codon_start=1	
		/product="AID"	
		/protein_id="AAM95403.1"	
		/db_xref="GI:22297220"	
		/translation="MDSLMMNRKFLYQKNVRWAKGRRETYLCVYVRRDSATSFSL	
		DFGLRNKNGCHVELLFURYISDWLDPRGCRYRVTFWTSRSPCYDCARHVADELGRNP	
		NLSRIETFARLYFCEDRKAEPGLRLRHAGVQIAITFEKDYFYCWNTFVENHRTFK	
		AWEGIHENSVRLSRLRLILLPLVEVDLDRDAFRTGL"	
BASE COUNT	127 a	165 c	155 g
ORIGIN	149 t		
Alignment Scores:			
Pred. No.:	1..51e-111	Length:	596
Score:	114.00	Matches:	197
Percent Similarity:	98.99%	Conservative:	0
Best Local Similarity:	98.99%	Mismatches:	1
Query Match:	57.58%	Indels:	2
DB:	9	Gaps:	0
US-09-966-880A-8 (1-198) x AF529816 (1-596)			
QY	1	MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp	20
DB	1	ATGGACAGCCTCTTGTGACACCGGAGGAGTTCCTTTACCAATTCAAAAATGTCGCTGG	60
QY	21	AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr	40
DB	61	GCTAAGGGTCGGCGTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGCACAGTCTACA	120
QY	41	SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu	60
DB	121	TCTTTTTCACGTGACTTTGGTTATCTTCGCAATAAGAACGCTGCCACGTGGAATGGCTC	180
QY	61	PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp	80
DB	181	TTCTCCCGCTACATCTCGGACTGGGACCTAGACCCCTGCCGCTGTACCGCTCACCTGG	240
QY	81	PheThrSerTrp-SerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGl	100
DB	241	TTACCTCCCG-GAGCCCTCTGTAGACTGTGCCGACATGTGCCGACTTTTCTCGGAGG	299
QY	100	YAsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgly	120
DB	300	GAACCCCAACCTCAGTCTGAGGATCTTACCCTGCCGCTCTACTTCTGTGAGGACCGCAA	359
QY	120	sAlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetTh	140
DB	360	GGCTGAGCCCGAGGGCTGCCGCGCTGCACCGCGGGGTGCACCAATAGCCATCATGAC	419
QY	140	rPheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPhely	160
DB	420	CTTCAAGATATTATTTTACTGCTGGAATACTTTGTGAAACCATGAAGAACCTTTCAA	479
QY	160	sAlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLe	180
DB	480	AGCCTGGGAGGGCTGCATGAAATTCAGTTCGTCTCTCCAGACAGCTTCGGCGCATCT	539
QY	180	uLeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThrLeuGlyLeu	198
DB	540	TTTGCCCTGTATGAGTTGATGACTTACGAGACGCAITTCGTACTTTTGGGACTT	594
RESULT 42			
AF529827			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source			
gene			

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (bases 1 to 596)  
Martin, A. and Scharff, M.D.

## AUTHORS

Somatic hypermutation of the AID transgene in B cells and non-B  
cells

## TITLE

Unpublished

## JOURNAL

2 (bases 1 to 596)

## REFERENCE

Martin, A. and Scharff, M.D.

## AUTHORS

Direct Submission

## TITLE

Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of  
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA

## JOURNAL

Location/Qualifiers

## FEATURES

source

1..596

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="Ramos 13"

1..>596

/gene="AID"

1..>596

/note="integrated into Burkitt's lymphoma cell line Ramos"

/codon\_start=1

/product="AID"

/protein\_id="AAM95414.1"

/db\_xref="GI:22297242"

/translation="MDSLMMNRKFLYOFKNVRWAKGRRETYLCYVVKRRDSATSFSL

DFGLRNKNGCHVELLFRLYISDMDLDPGRCYRVTWFTSWSPCYDCARHVDPLRGNP

NLSRIETARLYFCEDRKAEPGLRLRHAGVQTAIMTFKDYFCWNTFVENHERTEFK

AWEGHENSRLSRQLRLLPLVEVDLDRADFRLGL"

## BASE COUNT

127 a 163 c 155 g 151 t

## ORIGIN

## Alignment Scores:

Pred. No.: 3 1e-106 Length: 596

Score: 109.00 Matches: 197

Percent Similarity: 98.99% Conservative: 0

Best Local Similarity: 98.99% Mismatches: 1

Query Match: 55.05% Indels: 2

DB: 9 Gaps: 0

US-09-966-880A-8 (1-198) x AF529827 (1-596)

Qy 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20

Db 1 ATGGACAGCCTCTTATGATGAACGGAGGAAGTTCTTTACCAATTCAAAATGTCGCTGG 60

Qy 21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40

Db 61 GCTAAGGTCGCGTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGCTACA 120

Qy 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60

Db 121 TCCITTTTACCTGGACTTTGGTTATCTTCCGAATAAGACGCTGCCACGTGAATGCTC 180

Qy 61 PheLeuArgTyrTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80

Db 181 TTCCTCCGCTACATCTCGACTGGGACCTAGACCTGACCGCTGCTACCGCTCACCTGG 240

Qy 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100

Db 241 TTCACCTCCTGGAGCCCTGCTAGACATGTCGCCGACATGCGCGACTTTCTCGCAGAGG 300

Qy 101 AsnProAsnLeuSerLeuArgIlePheThr-AlaArgLeuTyrPheCysGluAspArgLy 120

Db 301 AACCCCAACCTCAGTCTGAGATCTTCAT-CGCGCGCTCTACTTCTGTGAGGACCGCAA 359

Qy 120 sAlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetTh 140

Db 360 GCCTGAGCCCGAGGGGCTGCGCGCTGACCGCGCGGTGCAAAATAGCCATCATGAC 419

Qy 140 rPheLysAspTyrPheTyrCysTyrAsnThrPheValGluAsnHisGluArgThrPheLy 160

Db 420 CTTCAAGATTATTTTACTGCTGCAATCTTTCTAGAAAAACCATGAAGAATTTTCAA 479

Qy 160 sAlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180

Db 480 AGCCTGGGAAGGGCTGCATGAAATTCAGTTCTCTCCACAGACGCTTCGGCGATCCT 539

Qy 180 uLeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThrLeuGlyLeu 198

Db 540 TTTGCCCTCTGATGAGTTGATGACTTACGACAGCCATTTTCGTACTTTTGGACCTT 594

RESULT 43

AF529854

LOCUS

DEFINITION

Cricetulus griseus clone 14 transgenic Homo sapiens AID (AID) mRNA,

complete cds.

AF529854

VERSION

AF529854.1 GI:22297295

KEYWORDS

SOURCE

ORGANISM

Chinese hamster.

Cricetulus griseus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Cricetinae;

Cricetulus.

REFERENCE

1 (bases 1 to 597)

Martin, A. and Scharff, M.D.

Somatic hypermutation of the AID transgene in B cells and non-B

cells

Unpublished

REFERENCE

2 (bases 1 to 597)

Martin, A. and Scharff, M.D.

Direct Submission

JOURNAL

Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of

Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA

FEATURES

Location/Qualifiers

1..597

/organism="Cricetulus griseus"

/db\_xref="taxon:10029"

/clone="14"

/cell\_line="CHO"

/transgenic

1..597

/organism="Homo sapiens"

/db\_xref="taxon:9606"

1..597

/gene="AID"

1..597

/gene="AID"

/codon\_start=1

/product="AID"

/protein\_id="AAM95441.1"

/db\_xref="GI:22297296"

/translation="MDSLMMNRKFLYOFKNVRWAKGRRETYLCYVVKRRDSATSFSL

DFGLRNKNGCHVELLFRLYISDMDLDPGRCYRVTWFTSWSPCYDCARHVDPLRGNP

NLSRIETARLYFCEDRKAEPGLRLRHAGVQTAIMTFKDYFCWNTFVENHERTEFK

AWEGHENSRLSRQLRLLPLVEVDLDRADFRTWGR"

BASE COUNT

128 a 163 c 158 g 148 t

ORIGIN

Alignment Scores:

Pred. No.: 3 59e-105 Length: 597

Score: 108.00 Matches: 194

Percent Similarity: 98.98% Conservative: 0

Best Local Similarity: 98.98% Mismatches: 1

Query Match: 54.55% Indels: 2

DB: 12 Gaps: 0

US-09-966-880A-8 (1-198) x AF529854 (1-597)

Qy 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20

Db 1 ATGGACAGCCTCTTATGATGAACGGAGGAAGTTCTTTACCAATTCAAAATGTCGCTGG 60

Qy 21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40

```

Db      61  GCTAAGGTCGGCGTAGACCTACCTGTCTACGTAGTGAAGAGCGGTGACAGTGTCTACA 120
QY      41  SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
Db      121  TCCTTTTCACTGGACTTTGGTTATCTTCGCATAGAAGCGGTGCCACGTGAATGTCTC 180
QY      61  PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
Db      181  TTCTCTCGCTACATCTCGGACTCGGACCTAGACCTGTGGCGCTGTCTACCGCTCACCTGG 240
QY      81  PheThrSerTrpSer-ProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgL 100
Db      241  TTCACCTCTCGAGCCCTGG-TAGACTGTGCCGACATGTGGCGGACTTTCTGCGAGG 299
QY      100  YAsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgly 120
Db      300  GAACCCCAACCTCAGTCTGAGGATCTTACCAGCGGCGCTCTACTTCTGTGAGGACCGCAA 359
QY      120  sAlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMeth 140
Db      360  GGCTGAGCCGAGGGGCTCGCGGCTGCACCGCGCGGGGTGCAATAGCCATCATGAC 419
QY      140  rPheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLy 160
Db      420  CTTCAAGATATTTTACTGCTGGAATACTTTTGTAGAAACCATGAAGAAGTTCATCA 479
QY      160  sAlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIle 180
Db      480  AGCCTGGGAAGGCTGCATGAATTCAGTTCGTCTCTCCACACAGCTTCGGCGCATCTC 539
QY      180  uLeuProLeuTyrGluValAspLeuArgAspAlaPheArgThr 195
Db      540  TTTGGCCCTGTATGAGTTGATGACTTACGAGACGCAATTCGTACT 585

RESULT 44
AF529817
LOCUS      596 bp mRNA linear PRI 19-AUG-2002
DEFINITION Homo sapiens clone Ramos 3 AID (AID) mRNA, partial cds.
ACCESSION AF529817
VERSION AF529817.1 GI:22297221
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B
        cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
        Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
    source
        1..596
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="Ramos 3"
            1..>596
            /gene="AID"
            1..>596
            /gene="AID"
            /note="Integrated into Burkitt's lymphoma cell line Ramos"
            /codon_start=1
            /product="AID"
            /protein_id="AAM95404.1"
            /db_xref="GI:22297222"
            /translation="MDSLIMNRKFLYOFKNVRWAKGRRTYLCYVVKRRDSATSFSL
            DFGYLNRKNGCHVELFLRYISDWLDPGRCYRVTWFTSWSPCYDCARHVDILRGNP
            NLSLRFITARLYFCEDRKAEPGLRLHRAGVQIAIMTFKDDFYCWNTFVENHRTFK

```

```

BASE COUNT      128 a      164 c      156 g      148 t
ORIGIN
AWEGLUHENSVRLSROLRRILLPLYEVDDLDRAFRTLGL"

Alignment Scores:
Pred. No.:      2,03e-92      Length:      596
Score:          96.00      Matches:      196
Percent Similarity: 98.00%      Conservative: 0
Best Local Similarity: 98.00%      Mismatches: 2
Query Match:      48.48%      Indels:      4
DB:              9      Gaps:      0

US-09-966-880A-8 (1-198) x AF529817 (1-596)

QY      1  MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
Db      1  ATGGACAGCCTCTTCATGAACCGGAGGAAGTTCTTTACCAATTCAAAATCTCCGCTGG 60
QY      21  AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40
Db      61  GCTAAGGGTCGGCGTGAACCTTACCCTGTCTACGTAGTGAAGAGCGGTGACAGTGTCTACA 120
QY      41  SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
Db      121  TCCTTTTCACTGGAGCTTTGGTTATCTTCGCAATAAGACGGCTGCCACGTGGAATGTCTC 180
QY      61  PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
Db      181  TTCCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCCGCTGTACCGCTCACCTGG 240
QY      81  PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPhe-LeuArgL 100
Db      241  TTCACCTCTCTGGAGCCCTGTCTACGACTGTGCCCGCACATGTGCCCGACAT-TCTGCGAGG 299
QY      100  YAsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgly 120
Db      300  GAACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAA 359
QY      120  sAlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMeth 140
Db      360  GGCTGAGCCGAGGGGCTCGCGGCTGCACCGCGCGGGGTGCAATAGCCATCATGAC 419
QY      140  rPheLysAspTyr-PheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheL 160
Db      420  CTTCAAGATAGA-TTTTACTGCTGGAATACTTTGTAGAAACCATGAAGAAGTTCATCA 478
QY      160  ysAlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleL 180
Db      479  AAGCCTGGGAGGGCTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGGCGCATCC 538
QY      180  euLeuProLeuTyrGluValAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
Db      539  TTTTGGCCCTGTATGAGTTGATGACTTACGAGACGCAATTCGTACTTTGGGACIT 594

RESULT 45
AB040430
LOCUS      11204 bp DNA linear PRI 03-OCT-2000
DEFINITION Homo sapiens AID gene for activation-induced cytidine deaminase,
        complete cds.
ACCESSION AB040430
VERSION AB040430.1 GI:9988407
KEYWORDS AID; activation-induced cytidine deaminase.
SOURCE Homo sapiens DNA.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Muto,T., Muramatsu,M., Taniwaki,M., Kinoshita,K. and Honjo,T.
TITLE Isolation, tissue distribution, and chromosomal localization of the
        human activation-induced cytidine deaminase (AID) gene
JOURNAL Genomics 68 (1), 85-88 (2000)
MEDLINE 20408890
REFERENCE 2 (sites)

```

AUTHORS Revy,P., Muto,T., Levy,Y., Geissmann,F., Plebani,A., Sanal,O., Catalan,N., Forveille,M., Dufourcq-Lagelouse,R., Gennery,A., Tezcan,I., Ersey,F., Kayserili,H., Ugazio,A.G., Brousse,N., Muramatsu,M., Notarangelo,L.D., Kinoshita,K., Honjo,T., Fischer,A. and Durandy,A.

TITLE Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2)

JOURNAL Cell 102 (5), 565-575 (2000)

MEDLINE 20460541

REFERENCE 3 (bases 1 to 11204)

AUTHORS Muto,T., Muramatsu,M., Taniwaki,M., Kinoshita,K. and Honjo,T.

TITLE Direct Submission

JOURNAL Submitted (18-MAR-2000) Tasuku Honjo, Kyoto University, Department of Medical Chemistry, Faculty of Medicine; Yoshida, Sakyo-ku, Kyoto, Kyoto 606-8501, Japan (E-mail:honjoemfour.med.kyoto-u.ac.jp, Tel:81-75-753-4371(ex.4371), Fax:81-75-753-4388)

FEATURES

Source Location/Qualifiers

1..11204

/organism="Homo sapiens"

/db\_xref="taxon:9606"

Join(521..528,6280..6427,7807..8077,8371..8486,8956..9009)

/gene="AID"

Join(521..528,6280..6427,7807..8077,8371..8486,8956..9009)

/gene="AID"

/codon\_start=1

/product="activation-induced cytidine deaminase"

/protein\_id="BAB12720.1"

/db\_xref="GI:9988408"

/translation="MDSLMMNRKFLYQKKNWAKGRREYLCVVKRRDSATSFSLDFGLRNKNGCVHVELLFURYISDMLDPRGCRYVTWTFSPPCDCAHVAADFLRGNP NLSLRIFARLYFCEDRKAEPGLRLHRAVGQTAIMTFKDFYCNWTFVENHRTFK AWEGLHENSVRSLRLLPLLEVDLDAFRFLGL"

BASE COUNT 3305 a 2273 c 2373 g 3253 t

ORIGIN

Alignment Scores:

Pred. No.: 6,1e-85 Length: 11204

Score: 90.00 Matches: 90

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 45.45% Indels: 0

DB: 9 Gaps: 0

US-09-966-880A-8 (1-198) x AB040430 (1-11204)

QY 53 AsnGlyCysHisValgluLeuPheLeuArgTyrIleSerAspTrpAspLeuAspPro 72

DB 7807 AACGGCTGCCAGTGGAAATGCTCTCTCCGCTACATCTCGGACTGGGACCTAGACCT 7866

QY 73 GlyArgCysTyrArgValThrTrpPheThrSerTrpSerProCysTyrAspCysAlaArg 92

DB 7867 GCCCGCTGCTACCGCGTACCTGGTTCACCTCTCGAGCCCTCTAGACTGTGCCCGA 7926

QY 93 HisValAlaAspPheLeuArgGlyAsnProAsnLeuSerLeuArgIlePheThrAlaArg 112

DB 7927 CATGTGGCCGACTTCTCGAGGAGAACCCCAACCTCAGTCTGAGATCTTCACCGCGGC 7986

QY 113 LeuTyrPheCysGluAspArgLysAlaGluProGluGlyLeuArgArgLeuHisArgAla 132

DB 7987 CTCCTACTTCTGTAGAGCCGCAAGGCTGAGCCGAGGGCTGCGCGCTGCACCGGCC 8046

QY 133 GlyValGlnIleAlaIleMetThrPheLys 142

DB 8047 GGGGTGCAATAGCATCATGACCTTCAAA 8076

RESULT 46

AC092184

LOCUS AC092184 71132 bp DNA linear PRI 12-JUN-2002

DEFINITION Homo sapiens 12 BAC RP11-438L7 (Roswell Park Cancer Institute Human BAC library) complete sequence.

ACCESSION AC092184 AC013443

VERSION AC092184.7 GI:21206067

KEYWORDS HTG.

## SOURCE ORGANISM

Homo sapiens

## REFERENCE AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

1 (bases 1 to 71132)

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amarantunga,H.C., Are,J.R., Ayele,M., Banks,T., Barbaria,J., Benton,J., Binage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Delaney,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaneye,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Emerling,S., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K., Hart,M., Haviak,P., Howe,A., Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C., Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Toshikhes,I., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louised,H., Lozado,R.J., Lu,X., Luciet,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Marondel,I., Martin,R., Martindale,A., Martinez,E., Massey,E., Mahoney,E., McLeod,M.P., Meador,M., Mel,G., Merscher,S., Metzker,M., Miller,A., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,M., Morris,S., Moser,M., Neal,D., Nelson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokkwo,S., Oguh,M., Okunolu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Roife,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shim,C., Shoohtari,N., Sisson,I., Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watling,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Kuchelapati,R., Weinstock,G. and Gibbs,R.

Direct Submission

Unpublished

2 (bases 1 to 71132)

Worley,K.C.

Direct Submission

Submitted (25-JUN-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 71132)

Worley,K.C.

Direct Submission

Submitted (18-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

4 (bases 1 to 71132)

Worley,K.C.

Direct Submission

Submitted (25-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

5 (bases 1 to 71132)

Worley,K.C.

Direct Submission

Submitted (12-JUN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

## COMMENT

On May 25, 2002 this sequence version replaced gi:20901754.  
 INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email  
[gc-help@bcm.tmc.edu](mailto:gc-help@bcm.tmc.edu)

CLONE LENGTH: This sequence does not necessarily represent the  
 entire insert of this clone. Overlapping regions of clones are only  
 sequenced and submitted once, so the sequence for the remainder of  
 the insert may be found in the record for the adjacent clones.  
 Overlapping clones are noted at the beginning and end of the  
 Features listing.

## ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches  
 of a local database that includes entries from dbSTS, GDB, and  
 local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green,  
 unpublished.) for Human and Mouse sequences.

Genes and region of sequence similarity are identified by BLAST  
 (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the  
 EST and cDNA sequences. Genes demonstrate at least two exons  
 flanked by consensus splice sites that maintained sequence  
 continuity across the splice junctions. Sequences that are not  
 identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum  
 standard of double strand coverage with a minimum of 2 clones and 2  
 reads with no ambiguities or 2 chemistries with a minimum of 2  
 clones and 3 reads with no ambiguities. If the sequence quality for  
 a region does not meet this standard, it will be indicated in the  
 annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality  
 standards - estimated error rate less than 1 per 10,000 bases.  
 Reports of lowest quality individual bases and measures of base  
 quality are listed below. Description of the metrics can be found  
 at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

## QUALSTAT-REPORT.

FEATURES	Location/Qualifiers
Source	1..71132
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="12"
misc_feature	/clone="RP11-438L7"
	complement(1..1448)
	/note="overlaps bases 1..1448 of clone AC092490"
	/function="clone overlap"
STS	187..408
	/standard_name="57233"
STS	439..560
	/standard_name="92005"
repeat_region	774..881
	/rpt_family="MIR"
repeat_region	903..1190
	/rpt_family="AluSq"
repeat_region	1191..1213
	/rpt_family="AT_rich"
STS	1591..1807
	/standard_name="6612"
STS	1744..1819
	/standard_name="8198"
repeat_region	1966..2264
	/rpt_family="AluSq"
repeat_region	3296..3328
	/rpt_family="(TTTC)n"
repeat_region	4097..4249
	/rpt_family="AluSq"
repeat_region	4360..4654
	/rpt_family="AluY"
repeat_region	complement(5140..5262)
	/rpt_family="FLAM_C"
repeat_region	5275..5304

repeat_region	/rpt_family="AT_rich"	3.01e-84	Length:	71132
5305..5563	/rpt_family="AluJo"	90.00	Matches:	90
5564..5599	/rpt_family="AT_rich"	100.00%	Conservative:	0
5644..5733	/rpt_family="MSTD"	100.00%	Mismatches:	0
5734..6012	/rpt_family="AluSq"	45.45%	Indels:	0
6020..6336	/rpt_family="AluX"	9	Gaps:	0
6337..6629	/rpt_family="MSTD"			
6630..6650	/rpt_family="(CAA)n"			
complement(6711..6817)	/rpt_family="MIR"			
complement(7033..7327)	/rpt_family="AluX"			
7607..7745	/rpt_family="FLAM_A"			
7980..8047	/rpt_family="MIR"			
complement(8050..8549)	/rpt_family="LTR47A"			
8567..8698	/rpt_family="MIR"			
complement(8815..9101)	/rpt_family="AluSg"			
complement(11227..12175)	/rpt_family="LTR5"			
STS	12625..12900			
	/standard_name="87432"			
repeat_region	complement(13547..13664)			
repeat_region	complement(13665..13979)			
repeat_region	complement(13980..14193)			
repeat_region	14622..14924			
repeat_region	/rpt_family="AluY"			
repeat_region	complement(15813..17136)			
repeat_region	complement(17272..17395)			
repeat_region	17426..17575			
repeat_region	/rpt_family="MIR"			
repeat_region	19021..19314			

Alignment Scores:  
 Pred. No.: 3.01e-84 Length: 71132  
 Score: 90.00 Matches: 90  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 45.45% Indels: 0  
 DB: 9 Gaps: 0

US-09-966-880A-8 (1-198) x AC092184 (1-71132)

QY	53	AsnGlyCysHisValGluLeuLeuPheLeuArgTyrIleSerAspTrpAspLeuAspPro	72
Db	42351	AAACGGCTGCACGTGGAAATGGCTCTCTCCGCTACATCTCGACTGGGACCTAGACCCT	42410
QY	73	GlyArgCysTyrArgValThrTrpPheThrSerTrpSerProCysTyrAspCysAlaAArg	92
Db	42411	GGCCCGCTGCTACCCGCTCACCTGGTTCACCTCTGGAGCCCTGTACGACGTGCCCCGA	42470
QY	93	HisValAlaAspPheLeuArgGlyAsnProAsnLeuSerLeuArgIlePheThrAlaAArg	112
Db	42471	CATGTGGCGCGACTTTCTCGAGGGAGACCCACCTCAGTCTGAGGATCTTCACCGCGCG	42530
QY	113	LeuTyrPheCysGluAspArgLysAlaGluProGluGlyLeuArgLysLeuHisArgAla	132

Db 42531 CTCTACTTCTGTGAGACCGCAAGGCTGAGCCCGAGGGGCTGCGCGGCTGCACCGGCC 42590

QY 133 GlyValGlnIleAlaIleMetThrPheLys 142

Db 42591 GGGGTGCAATAGCATCATGACCTTCARA 42620

RESULT 47

AF132979

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

gene

CDS

BASE COUNT

ORIGIN

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

US-09-966-880A-8 (1-198) x AF132979 (1-2440)

QY 54 GlyCysHisValGluLeuPheLeuArgTyrIleSerAspTrpAspLeuAspProGly 73

Db 252 GGCTGCCAGTGAATTGTTCTACGCTACATCTCAGACTGGAGCTGGACCCGGC 311

QY 74 ArgCysTyrArgValThrTrpPheThrSerTrpSerProCysTyrAspCysAlaArgHis 93

Db 312 CGGTGTTACCGGCTCACCTGGTTACCTCCCTGGAGCCGCTGCTATGACTGTGCCGGCAC 371

QY 94 ValAla 95

Db 372 GTGGCT 377

RESULT 48

AC094826

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

gene

CDS

BASE COUNT

ORIGIN

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

US-09-966-880A-8 (1-198) x AF132979 (1-2440)

QY 54 GlyCysHisValGluLeuPheLeuArgTyrIleSerAspTrpAspLeuAspProGly 73

Db 252 GGCTGCCAGTGAATTGTTCTACGCTACATCTCAGACTGGAGCTGGACCCGGC 311

QY 74 ArgCysTyrArgValThrTrpPheThrSerTrpSerProCysTyrAspCysAlaArgHis 93

Db 312 CGGTGTTACCGGCTCACCTGGTTACCTCCCTGGAGCCGCTGCTATGACTGTGCCGGCAC 371

QY 94 ValAla 95

Db 372 GTGGCT 377

AC094826 143044 bp DNA linear HTG 20-DEC-2001  
Rattus norvegicus clone CH230-2013, \*\*\* SEQUENCING IN PROGRESS \*\*\*,  
61 unordered pieces.

AC094826

AC094826.2 GI:17941606

HTG: HTGS-PHASE1

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

gene

CDS

BASE COUNT

ORIGIN

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

US-09-966-880A-8 (1-198) x AF132979 (1-2440)

QY 54 GlyCysHisValGluLeuPheLeuArgTyrIleSerAspTrpAspLeuAspProGly 73

Db 252 GGCTGCCAGTGAATTGTTCTACGCTACATCTCAGACTGGAGCTGGACCCGGC 311

QY 74 ArgCysTyrArgValThrTrpPheThrSerTrpSerProCysTyrAspCysAlaArgHis 93

Db 312 CGGTGTTACCGGCTCACCTGGTTACCTCCCTGGAGCCGCTGCTATGACTGTGCCGGCAC 371

QY 94 ValAla 95

Db 372 GTGGCT 377

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

Project Information

Center project name: GBMC

Center clone name: CH230-2013

Summary Statistics

Program: Phrap: version 0.990329First call to

On Dec 20, 2001 this sequence version replaced gi:15424662.

Submitted (15-SEP-2001) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

Project Information

Center project name: GBMC

Center clone name: CH230-2013

Summary Statistics

Program: Phrap: version 0.990329First call to

On Dec 20, 2001 this sequence version replaced gi:15424662.

Submitted (15-SEP-2001) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

Project Information

Center project name: GBMC

Center clone name: CH230-2013



## findPhrapList

Consensus quality: 110311 bases at least Q40  
 Consensus quality: 118777 bases at least Q30  
 Consensus quality: 125541 bases at least Q20  
 Estimated insert size: 109318; sum-of-contigs estimation  
 Quality coverage: 0x in Q20 bases; agarose-1p estimation  
 Quality coverage: 1.7x in Q20 bases; sum-of-contigs estimation  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 61 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1  
 \* 8004 8103: contig of 8003 bp in length  
 \* 8104 8103: gap of unknown length  
 \* 12050 12049: contig of 3946 bp in length  
 \* 12150 12149: gap of unknown length  
 \* 16675 16674: contig of 4525 bp in length  
 \* 16775 16774: gap of unknown length  
 \* 16775 22570: contig of 5796 bp in length  
 \* 22571 22570: gap of unknown length  
 \* 22671 27595: contig of 4925 bp in length  
 \* 27596 27595: gap of unknown length  
 \* 27696 31790: contig of 4095 bp in length  
 \* 31791 31890: gap of unknown length  
 \* 31891 35625: contig of 3735 bp in length  
 \* 35626 35725: gap of unknown length  
 \* 35726 38511: contig of 2786 bp in length  
 \* 38511 38611: gap of unknown length  
 \* 38612 41020: contig of 2409 bp in length  
 \* 41021 41120: gap of unknown length  
 \* 41121 44145: contig of 3025 bp in length  
 \* 44146 44245: gap of unknown length  
 \* 44246 46975: contig of 2730 bp in length  
 \* 46976 47075: gap of unknown length  
 \* 47076 50825: contig of 3750 bp in length  
 \* 50826 50925: gap of unknown length  
 \* 50926 55401: contig of 4476 bp in length  
 \* 55402 55501: gap of unknown length  
 \* 55502 58143: contig of 2642 bp in length  
 \* 58144 58243: gap of unknown length  
 \* 58244 60960: contig of 2617 bp in length  
 \* 60961 60960: gap of unknown length  
 \* 60961 63226: contig of 2266 bp in length  
 \* 63227 63326: gap of unknown length  
 \* 63327 66558: contig of 3232 bp in length  
 \* 66559 69476: gap of unknown length  
 \* 69477 69576: gap of unknown length  
 \* 69577 72308: contig of 2732 bp in length  
 \* 72309 72408: gap of unknown length  
 \* 72409 74936: contig of 2528 bp in length  
 \* 74937 75036: gap of unknown length  
 \* 75037 77300: contig of 2264 bp in length  
 \* 77301 77400: gap of unknown length  
 \* 77401 80440: contig of 3040 bp in length  
 \* 80441 80540: gap of unknown length  
 \* 80541 83286: contig of 2746 bp in length  
 \* 83287 83387: gap of unknown length  
 \* 83387 85219: contig of 1833 bp in length  
 \* 85220 85319: gap of unknown length  
 \* 85320 87656: contig of 2337 bp in length  
 \* 87657 89703: contig of 1947 bp in length  
 \* 89704 89803: gap of unknown length  
 \* 89804 91300: contig of 1497 bp in length  
 \* 91301 91401: gap of unknown length  
 \* 92768: contig of 1368 bp in length

\* 92769 92868: gap of unknown length  
 \* 92869 94158: contig of 1290 bp in length  
 \* 94159 94258: gap of unknown length  
 \* 94259 96077: contig of 1819 bp in length  
 \* 96078 96177: gap of unknown length  
 \* 96178 97861: contig of 1684 bp in length  
 \* 97862 97961: gap of unknown length  
 \* 97962 99629: contig of 1668 bp in length  
 \* 99630 99729: gap of unknown length  
 \* 99730 101901: contig of 2172 bp in length  
 \* 101902 102001: gap of unknown length  
 \* 102002 103791: contig of 1790 bp in length  
 \* 103792 103891: gap of unknown length  
 \* 103892 105124: contig of 1233 bp in length  
 \* 105125 105224: gap of unknown length  
 \* 105225 106434: contig of 1210 bp in length  
 \* 106435 106535: gap of unknown length  
 \* 106535 108041: contig of 1507 bp in length  
 \* 108042 108141: gap of unknown length  
 \* 108142 109842: contig of 1701 bp in length  
 \* 109843 109942: gap of unknown length  
 \* 109943 112159: contig of 2217 bp in length  
 \* 112160 112259: gap of unknown length  
 \* 112260 113621: contig of 1362 bp in length  
 \* 113622 113721: gap of unknown length  
 \* 113722 115223: contig of 1502 bp in length  
 \* 115224 115323: gap of unknown length  
 \* 115324 116353: contig of 1030 bp in length  
 \* 116354 116453: gap of unknown length  
 \* 116454 118323: contig of 1870 bp in length  
 \* 118324 118423: gap of unknown length  
 \* 118424 119490: contig of 1067 bp in length  
 \* 119491 119590: gap of unknown length  
 \* 119591 120620: contig of 1030 bp in length  
 \* 120621 120720: gap of unknown length  
 \* 120721 122038: contig of 1318 bp in length  
 \* 122039 122138: gap of unknown length  
 \* 122139 123239: contig of 1101 bp in length  
 \* 123240 123339: gap of unknown length  
 \* 123340 124595: contig of 1256 bp in length  
 \* 124596 124695: gap of unknown length  
 \* 124696 125975: contig of 1280 bp in length  
 \* 125976 126075: gap of unknown length  
 \* 126076 127220: contig of 1145 bp in length  
 \* 127221 127320: gap of unknown length  
 \* 127321 128724: contig of 1404 bp in length  
 \* 128725 128824: gap of unknown length  
 \* 128825 130661: contig of 1837 bp in length  
 \* 130662 130761: gap of unknown length  
 \* 130762 132154: contig of 1393 bp in length  
 \* 132155 132254: gap of unknown length  
 \* 132255 133282: contig of 1028 bp in length  
 \* 133283 133382: gap of unknown length  
 \* 133383 134875: contig of 1493 bp in length  
 \* 134876 134975: gap of unknown length  
 \* 134976 136074: contig of 1099 bp in length  
 \* 136075 136174: gap of unknown length

## Alignment Scores:

Pred. No.: 5.67e-33 Length: 143044  
 Score: 42.00 Matches: 42  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 21.21% Indels: 0  
 DB: 2 Gaps: 0

US-09-966-880A-8 (1-198) x AC094826 (1-143044)

Qy 54 GlyCysHisValGluLeuLeuPheLeuArgThrIleSerAspTrpAspLeuAspProGly 73  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 10093 GGCTGCCACGTGGAAATTGTTCTTCACGTACATCTCGGACCTGGACCCCGC 10152  
 Qy 74 ArgCysTyrArgValThrTrpPheThrSerTrpSerProCysTyrAspCysAlaArgHis 93



* 34414	36200:	contig of 1787	bp in length
* 36201	36300:	gap of unknown	length
* 36301	37867:	contig of 1567	bp in length
* 37868	37967:	gap of unknown	length
* 37968	40374:	contig of 2407	bp in length
* 40375	40474:	gap of unknown	length
* 40475	42090:	contig of 1616	bp in length
* 42091	42190:	gap of unknown	length
* 42191	43912:	contig of 1722	bp in length
* 43913	44012:	gap of unknown	length
* 44013	45650:	contig of 1638	bp in length
* 45651	45750:	gap of unknown	length
* 45751	47922:	contig of 2172	bp in length
* 47923	48022:	gap of unknown	length
* 48023	50032:	contig of 2010	bp in length
* 50033	50132:	gap of unknown	length
* 50133	51867:	contig of 1735	bp in length
* 51868	51967:	gap of unknown	length
* 51968	54813:	contig of 2846	bp in length
* 54814	54913:	gap of unknown	length
* 54914	56337:	contig of 1424	bp in length
* 56338	56437:	gap of unknown	length
* 56438	59133:	contig of 2696	bp in length
* 59134	59233:	gap of unknown	length
* 59234	61478:	contig of 2245	bp in length
* 61479	61578:	gap of unknown	length
* 61579	64321:	contig of 2743	bp in length
* 64322	64421:	gap of unknown	length
* 64422	67033:	contig of 2612	bp in length
* 67034	67133:	gap of unknown	length
* 67134	69227:	contig of 2094	bp in length
* 69228	69327:	gap of unknown	length
* 69328	71941:	contig of 2614	bp in length
* 71942	72041:	gap of unknown	length
* 72042	76107:	contig of 4066	bp in length
* 76108	76207:	gap of unknown	length
* 76208	78643:	contig of 2436	bp in length
* 78644	78743:	gap of unknown	length
* 78744	82956:	contig of 4213	bp in length
* 82957	83056:	gap of unknown	length
* 83057	86461:	contig of 3405	bp in length
* 86462	86561:	gap of unknown	length
* 86562	89125:	contig of 2564	bp in length
* 89126	89225:	gap of unknown	length
* 89226	91743:	contig of 2518	bp in length
* 91744	91843:	gap of unknown	length
* 91844	97444:	contig of 5601	bp in length
* 97445	97544:	gap of unknown	length
* 97545	101038:	contig of 3514	bp in length
* 101039	101158:	gap of unknown	length
* 101159	104772:	contig of 3614	bp in length
* 104773	104872:	gap of unknown	length
* 104873	109395:	contig of 4523	bp in length
* 109396	109495:	gap of unknown	length
* 109496	111919:	contig of 2424	bp in length
* 111920	111920:	gap of unknown	length
* 112020	119114:	contig of 7095	bp in length
* 119115	119214:	gap of unknown	length
* 119215	122620:	contig of 3406	bp in length
* 122621	122720:	gap of unknown	length
* 122721	128279:	contig of 5559	bp in length
* 128280	128379:	gap of unknown	length
* 128380	132063:	contig of 3684	bp in length

Alignment Scores:

Pred. No.:	6.82e-33	Length:	177103
Score:	42.00	Matches:	42
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	21.21%	Indels:	0
DB:	2	Gaps:	0

US-09-966-880A-8 (1-198) x AC120617 (1-177103)

QY	54	GlyCysHisValIclulLeuPheLeuArgTyrIleSerAspTrpAspLeuAspProGly	73
Db	142302	GGCTGCCACGTGAATTTGTTCTACGCTACATCTCGAGCTGGACCTGGACCCGGC	142361
QY	74	ArgCysTyrArgValThrTrpPheThrSerProCysTyrAspCysAlaArgHis	93
Db	142362	CGGTGTTACCGTGTCACTGGTTCACCTCCTGAGCCCTGCTACGACTGTGCGGCAC	142421
QY	94	ValAla	95
Db	142422	GTGGCT	142427

RESULT 50

AC119975	50269 bp	DNA	linear	HTG 02-MAY-2002
Mus musculus	clone RP24-483K3	LOW-PASS	SEQUENCE SAMPLING.	
AC119975				
DEFINITION	AC119975.1	GI:20389635		
ACCESSION	HTG; HTGS_PHASE0.			
VERSION				
KEYWORDS				
SOURCE	Mus musculus			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
TITLE	Birren,B., Linton,L., Nusbaum,C. and Lander,E.			
JOURNAL	Mus musculus, clone RP24-483K3			
REFERENCE	Unpublished			
AUTHORS	2 (bases 1 to 50269)			

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,X., Colangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., MacDonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE

Submitted (02-MAY-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

JOURNAL

COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L25744  
Center clone name: 483\_K\_3  
-----

- \* NOTE: This record contains 65 individual
- \* sequencing reads that have not been assembled into
- \* contigs. Runs of N are used to separate the reads
- \* and the order in which they appear is completely
- \* arbitrary. Low-pass sequence sampling is useful for
- \* identifying clones that may be gene-rich and allows
- \* overlap relationships among clones to be deduced.

\* However, it should not be assumed that this clone  
 \* will be sequenced to completion. In the event that  
 \* the record is updated, the accession number will  
 \* be preserved.

1  
 \* 685 784: contig of 684 bp in length  
 \* 785 1439: contig of 655 bp in length  
 \* 1440 1539: gap of 100 bp  
 \* 1540 2214: contig of 675 bp in length  
 \* 2215 2314: gap of 100 bp  
 \* 2315 3000: contig of 686 bp in length  
 \* 3001 3100: gap of 100 bp  
 \* 3101 3794: contig of 694 bp in length  
 \* 3795 3894: gap of 100 bp  
 \* 3895 4576: contig of 682 bp in length  
 \* 4577 4676: gap of 100 bp  
 \* 4677 5361: contig of 685 bp in length  
 \* 5362 5461: gap of 100 bp  
 \* 5462 6138: contig of 677 bp in length  
 \* 6139 6238: gap of 100 bp  
 \* 6239 6900: contig of 662 bp in length  
 \* 6901 7000: gap of 100 bp  
 \* 7001 7674: contig of 674 bp in length  
 \* 7675 7774: gap of 100 bp  
 \* 7775 8446: contig of 672 bp in length  
 \* 8447 8546: gap of 100 bp  
 \* 8547 9219: contig of 673 bp in length  
 \* 9220 9319: gap of 100 bp  
 \* 9320 9962: contig of 643 bp in length  
 \* 9963 10062: gap of 100 bp  
 \* 10063 10738: contig of 676 bp in length  
 \* 10739 10838: gap of 100 bp  
 \* 10839 11520: contig of 682 bp in length  
 \* 11521 11620: gap of 100 bp  
 \* 11621 12311: contig of 691 bp in length  
 \* 12312 12411: gap of 100 bp  
 \* 12412 13105: contig of 694 bp in length  
 \* 13106 13205: gap of 100 bp  
 \* 13206 13865: contig of 660 bp in length  
 \* 13866 13965: gap of 100 bp  
 \* 13966 14653: contig of 688 bp in length  
 \* 14654 14753: gap of 100 bp  
 \* 14754 15411: contig of 658 bp in length  
 \* 15412 15511: gap of 100 bp  
 \* 15512 16182: contig of 671 bp in length  
 \* 16183 16282: gap of 100 bp  
 \* 16283 16971: contig of 689 bp in length  
 \* 16972 17071: gap of 100 bp  
 \* 17072 17754: contig of 683 bp in length  
 \* 17755 17854: gap of 100 bp  
 \* 17855 18535: contig of 681 bp in length  
 \* 18536 18635: gap of 100 bp  
 \* 18636 19317: contig of 682 bp in length  
 \* 19318 19417: gap of 100 bp  
 \* 19418 20096: contig of 679 bp in length  
 \* 20097 20196: gap of 100 bp  
 \* 20197 20879: contig of 683 bp in length  
 \* 20880 20979: gap of 100 bp  
 \* 20980 21667: contig of 688 bp in length  
 \* 21668 21767: gap of 100 bp  
 \* 21768 22437: contig of 670 bp in length  
 \* 22438 22537: gap of 100 bp  
 \* 22538 23197: contig of 660 bp in length  
 \* 23198 23297: gap of 100 bp  
 \* 23298 23974: contig of 677 bp in length  
 \* 23975 24074: gap of 100 bp  
 \* 24075 24745: contig of 671 bp in length  
 \* 24746 24845: gap of 100 bp  
 \* 24846 25522: contig of 677 bp in length  
 \* 25523 25622: gap of 100 bp  
 \* 25623 26294: contig of 672 bp in length  
 \* 26295 26394: gap of 100 bp  
 \* 26395 27061: contig of 667 bp in length

\* 27062 27161: gap of 100 bp  
 \* 27162 27827: contig of 666 bp in length  
 \* 27828 27927: gap of 100 bp  
 \* 27928 28611: contig of 684 bp in length  
 \* 28612 28711: gap of 100 bp  
 \* 28712 29395: contig of 684 bp in length  
 \* 29396 29495: gap of 100 bp  
 \* 29496 30166: contig of 671 bp in length  
 \* 30167 30266: gap of 100 bp  
 \* 30267 30946: contig of 680 bp in length  
 \* 30947 31046: gap of 100 bp  
 \* 31047 31700: contig of 654 bp in length  
 \* 31701 31800: gap of 100 bp  
 \* 31801 32455: contig of 655 bp in length  
 \* 32456 32555: gap of 100 bp  
 \* 32556 33224: contig of 669 bp in length  
 \* 33225 33324: gap of 100 bp  
 \* 33325 34012: contig of 688 bp in length  
 \* 34013 34112: gap of 100 bp  
 \* 34113 34756: contig of 644 bp in length  
 \* 34757 34856: gap of 100 bp  
 \* 34857 35514: contig of 658 bp in length  
 \* 35515 35614: gap of 100 bp  
 \* 35615 36297: contig of 683 bp in length  
 \* 36298 36397: gap of 100 bp  
 \* 36398 37066: contig of 669 bp in length  
 \* 37067 37166: gap of 100 bp  
 \* 37167 37848: contig of 682 bp in length  
 \* 37849 37948: gap of 100 bp  
 \* 37949 38606: contig of 658 bp in length  
 \* 38607 38706: gap of 100 bp  
 \* 38707 39385: contig of 679 bp in length  
 \* 39386 39485: gap of 100 bp  
 \* 39486 40178: contig of 693 bp in length  
 \* 40179 40278: gap of 100 bp  
 \* 40279 40935: contig of 657 bp in length  
 \* 40936 41035: gap of 100 bp  
 \* 41036 41718: contig of 683 bp in length  
 \* 41719 41818: gap of 100 bp  
 \* 41819 42492: contig of 674 bp in length  
 \* 42493 42592: gap of 100 bp  
 \* 42593 43275: contig of 683 bp in length  
 \* 43276 43375: gap of 100 bp  
 \* 43376 44038: contig of 663 bp in length  
 \* 44039 44138: gap of 100 bp  
 \* 44139 44821: contig of 683 bp in length  
 \* 44822 44921: gap of 100 bp  
 \* 44922 45602: contig of 681 bp in length  
 \* 45603 45702: gap of 100 bp  
 \* 45703 46382: contig of 680 bp in length  
 \* 46383 46482: gap of 100 bp  
 \* 46483 47153: contig of 671 bp in length  
 \* 47154 47253: gap of 100 bp  
 \* 47254 47933: contig of 680 bp in length  
 \* 47934 48033: gap of 100 bp  
 \* 48034 48705: contig of 672 bp in length  
 \* 48706 48805: gap of 100 bp  
 \* 48806 49481: contig of 676 bp in length  
 \* 49482 49581: gap of 100 bp  
 \* 49582 50269: contig of 688 bp in length.

FEATURES

Source

1. .50269  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="RP24-483K3"

BASE COUNT 12034 a 9403 c 9549 g 12707 t 6576 others  
 ORIGIN

Alignment Scores:

Pred. No.: 4.09e-29 Length: 50269  
 Score: 38.00 Matches: 38  
 Percent Similarity: 100.00% Conservative: 0

Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	19.19%	Indels:	0
DB:	2	Gaps:	0
US-09-966-880A-8 (1-198) x AC119975 (1-50269)			
Qy	58	GluLeuLeuPheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCystYrArg 77	
Db	528	GAATTGTTGTCCTACGCTACATCTCAGACTGGACCTGCACCGCGCGGTGTTACCGC 587	
Qy	78	ValThrTrpPheThrSerProCysTyrAspCysAlaAargHisValIla 95	
Db	588	GTCACCTGGTTACCTCTCGGAGCCGCTGCTATGACTGCCCCGACGTGGCT 641	
RESULT 51			
AC109119/c			
LOCUS			
DEFINITION	Rattus norvegicus clone CH230-98A4, *** SEQUENCING IN PROGRESS ***, 46 unordered pieces.		
ACCESSION	AC109119		
VERSION	AC109119.3	GI:21738722	
KEYWORDS	HTG: HTGS_PHASE1.		
SOURCE	Norway rat.		
ORGANISM	Rattus norvegicus		
	Euteleostomi; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
REFERENCE	1 (bases 1 to 98161)		
AUTHORS	Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C., Alsbrooks, S.L., Anaratunga, H.C., Are, J.R., Ayele, M., Banks, T., Barbara, J., Benton, J., Bimaye, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buha, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carrott, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M., Hollway, C., Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseged, H., Lozaro, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoostari, N., Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, J., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.		
	Direct Submission		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 98161)		
AUTHORS	Worley, K.C.		
TITLE	Direct Submission		
JOURNAL	Submitted (03-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA		
AUTHORS	3 (bases 1 to 98161)		
TITLE	Worley, K.C.		
JOURNAL	Direct Submission		
COMMENT	Submitted (13-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA On Jul 12, 2002 this sequence version replaced gi:18860460. ----- Genome Center Center: Baylor College of Medicine Center code: BCM Web site: http://www.hgsc.bcm.tmc.edu/ Contact: hgsc-help@bcm.tmc.edu ----- Project Information Center project name: GREZ Center clone name: CH230-98A4 ----- Summary Statistics Sequencing vector: Plasmid; Chemistry: Dye-terminator Big Dye; 100% of reads Assembly program: Phrap; version 0.990329 Consensus quality: 54904 bases at least Q40 Consensus quality: 57914 bases at least Q30 Consensus quality: 60850 bases at least Q20 ----- * NOTE: Estimated insert size may differ from sequence length * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html). * NOTE: This is a 'working draft' sequence. It currently * consists of 46 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved. * 1 1023: contig of 1023 bp in length * 1123: gap of unknown length * 1124 2573: contig of 1450 bp in length * 2574 2673: gap of unknown length * 2674 4111: contig of 1438 bp in length * 4112 4211: gap of unknown length * 4211 5495: contig of 1284 bp in length * 5496 5595: gap of unknown length * 5596 6676: contig of 1081 bp in length * 6677 7884: contig of 1108 bp in length * 7885 7984: gap of unknown length * 7985 9198: contig of 1214 bp in length * 9199 9299: contig of 1161 bp in length * 9299 10459: gap of unknown length * 10460 11982: contig of 1423 bp in length * 11983 12082: gap of unknown length * 12083 13580: contig of 1498 bp in length * 13581 14896: contig of 1216 bp in length * 14897 16060: contig of 1064 bp in length * 16061 16160: gap of unknown length * 16161 17447: contig of 1287 bp in length * 17448 18812: contig of 1265 bp in length * 18813 18912: gap of unknown length * 18913 20137: contig of 1225 bp in length * 20138 20237: gap of unknown length * 20238 21396: contig of 1159 bp in length * 21397 21496: gap of unknown length * 21497 22738: contig of 1242 bp in length * 22739 22838: gap of unknown length * 22839 24155: contig of 1317 bp in length * 24156 24256: gap of unknown length * 24256 26129: contig of 1874 bp in length		

\* 26130 26229: gap of unknown length  
 \* 26230 27842: contig of 1613 bp in length  
 \* 27843 27942: gap of unknown length  
 \* 27943 29436: contig of 1494 bp in length  
 \* 29437 29536: gap of unknown length  
 \* 29537 31686: contig of 2150 bp in length  
 \* 31687 31786: gap of unknown length  
 \* 31787 33456: contig of 1670 bp in length  
 \* 33457 33556: gap of unknown length  
 \* 33557 34636: contig of 1080 bp in length  
 \* 34637 34736: gap of unknown length  
 \* 34737 37301: contig of 2565 bp in length  
 \* 37302 37401: gap of unknown length  
 \* 37402 38699: contig of 1298 bp in length  
 \* 38700 38799: gap of unknown length  
 \* 38800 40740: contig of 1941 bp in length  
 \* 40741 40840: gap of unknown length  
 \* 40841 43152: contig of 2312 bp in length  
 \* 43153 43252: gap of unknown length  
 \* 43253 45781: contig of 2529 bp in length  
 \* 45782 45881: gap of unknown length  
 \* 45882 48371: contig of 2390 bp in length  
 \* 48372 48371: gap of unknown length  
 \* 48372 49668: contig of 1597 bp in length  
 \* 49669 50669: gap of unknown length  
 \* 50670 52282: contig of 2214 bp in length  
 \* 52283 52382: gap of unknown length  
 \* 52383 54977: contig of 2595 bp in length  
 \* 54978 55077: gap of unknown length  
 \* 55078 57554: contig of 2577 bp in length  
 \* 57555 57754: gap of unknown length  
 \* 57755 59870: contig of 2116 bp in length  
 \* 59871 59970: gap of unknown length  
 \* 59971 62400: contig of 2430 bp in length  
 \* 62401 62500: gap of unknown length  
 \* 62501 64476: contig of 1976 bp in length  
 \* 64477 64577: gap of unknown length  
 \* 64577 67748: contig of 3172 bp in length  
 \* 67749 67848: gap of unknown length  
 \* 67849 70585: contig of 2737 bp in length  
 \* 70586 70686: gap of unknown length  
 \* 70687 73046: contig of 2361 bp in length  
 \* 73047 73146: gap of unknown length  
 \* 73147 76312: contig of 3166 bp in length  
 \* 76313 76412: gap of unknown length  
 \* 76413 79506: contig of 3094 bp in length  
 \* 79507 79606: gap of unknown length  
 \* 79607 83285: contig of 3679 bp in length  
 \* 83286 83385: gap of unknown length  
 \* 83386 88188: contig of 4803 bp in length  
 \* 88189 88288: gap of unknown length  
 \* 88289 93249: contig of 4961 bp in length  
 \* 93250 93349: gap of unknown length  
 \* 93350 98161: contig of 4812 bp in length.

FEATURES

source  
 i. .98161  
 /organism="Rattus norvegicus"  
 /db\_xref="taxon:10116"  
 /clone="CH230-98A4"  
 BASE COUNT 25850 a 20396 c 20306 g 25999 t 5610 others  
 ORIGIN

Alignment Scores:

Pred. No.: 1.3e-24 Length: 98161  
 Score: 34.00 Matches: 34  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 17.17% Indels: 0  
 DB: 2 Gaps: 0

US-09-966-880A-8 (1-198) x AC109119 (1-98161)

QY 54 GlyCysHisValGluLeuLeuPheLeuArgTyrIleSerAspTrpAspLeuAspProGly 73

Db 4378 GGCTCCACGTGGAATTGTTCTACGCTACATCTCGGACTGGACCTGACCCCGGC 4319  
 QY 74 ArgCysTyrArgValThrTrpPheThrSerTrpSerProCys 87  
 Db 4318 CGGTGTTACCGTGTACCTGGTTCACCTTCCTCGAGCCCTGC 4277  
 RESULT 52  
 AC119975/c 50269 bp DNA linear HTG 02-MAY-2002  
 LOCUS Mus musculus clone RP24-483K3, LOW-PASS SEQUENCE SAMPLING.  
 DEFINITION AC119975  
 ACCESSION AC119975  
 VERSION AC119975.1 GI:20389635  
 HTG: HTGS\_PHASE0.  
 KEYWORDS Mus musculus.  
 SOURCE Mus musculus.  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
 1 (bases 1 to 50269)  
 Mus musculus, clone RP24-483K3  
 Unpublished  
 2 (bases 1 to 50269)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
 Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,I.,  
 Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chung,J.,  
 Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,  
 Cook,A., Cooke,P., DeAtellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
 Fato,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,  
 Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
 Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
 Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazars,R.,  
 Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,  
 MacLean,C., MacDonald,P., Major,J., Marquis,N., Matthews,C.,  
 McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Meneus,L.,  
 Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,  
 Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,  
 Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,  
 Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,  
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupbach,R.,  
 Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
 Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodorovic,H.,  
 Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H.,  
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (02-MAY-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www.seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L25744  
 Center clone name: 483\_K\_3  
 -----  
 \* NOTE: This record contains 65 individual  
 \* sequencing reads that have not been assembled into  
 \* contigs. Runs of N are used to separate the reads  
 \* and the order in which they appear is completely  
 \* arbitrary. Low-pass sequence sampling is useful for  
 \* identifying clones that may be gene-rich and allows  
 \* overlap relationships among clones to be deduced.  
 \* However, it should not be assumed that this clone  
 \* will be sequenced to completion. In the event that  
 \* the record is updated, the accession number will  
 \* be preserved.  
 \* 1 584: contig of 684 bp in length  
 \* 685 784: gap of 100 bp  
 \* \*

```

* 785 1439: contig of 655 bp in length
* 1440 1539: gap of 100 bp
* 1540 2214: contig of 675 bp in length
* 2215 2314: gap of 100 bp
* 2315 3000: contig of 686 bp in length
* 3001 3100: gap of 100 bp
* 3101 3794: contig of 694 bp in length
* 3795 3894: gap of 100 bp
* 3895 4576: contig of 682 bp in length
* 4577 4676: gap of 100 bp
* 4677 5361: contig of 685 bp in length
* 5362 5461: gap of 100 bp
* 5462 6138: contig of 677 bp in length
* 6139 6238: gap of 100 bp
* 6239 6900: contig of 662 bp in length
* 6901 7000: gap of 100 bp
* 7001 7674: contig of 674 bp in length
* 7675 7774: gap of 100 bp
* 7775 8446: contig of 672 bp in length
* 8447 8546: gap of 100 bp
* 8547 9219: contig of 673 bp in length
* 9220 9319: gap of 100 bp
* 9320 9962: contig of 643 bp in length
* 9963 10062: gap of 100 bp
* 10063 10738: contig of 676 bp in length
* 10739 10838: gap of 100 bp
* 10839 11520: contig of 682 bp in length
* 11521 11620: gap of 100 bp
* 11621 12311: contig of 691 bp in length
* 12312 12411: gap of 100 bp
* 12412 13105: contig of 694 bp in length
* 13106 13205: gap of 100 bp
* 13206 13865: contig of 660 bp in length
* 13866 13965: gap of 100 bp
* 13966 14653: contig of 688 bp in length
* 14654 14753: gap of 100 bp
* 14754 15411: contig of 658 bp in length
* 15412 15511: gap of 100 bp
* 15512 16182: contig of 671 bp in length
* 16183 16282: gap of 100 bp
* 16283 16971: contig of 689 bp in length
* 16972 17071: gap of 100 bp
* 17072 17754: contig of 683 bp in length
* 17755 17854: gap of 100 bp
* 17855 18535: contig of 681 bp in length
* 18536 18635: gap of 100 bp
* 18636 19317: contig of 682 bp in length
* 19318 19417: gap of 100 bp
* 19418 20096: contig of 679 bp in length
* 20097 20196: gap of 100 bp
* 20197 20879: contig of 683 bp in length
* 20880 20979: gap of 100 bp
* 20980 21667: contig of 688 bp in length
* 21668 21767: gap of 100 bp
* 21768 22437: contig of 670 bp in length
* 22438 22537: gap of 100 bp
* 22538 23197: contig of 660 bp in length
* 23198 23297: gap of 100 bp
* 23298 23974: contig of 677 bp in length
* 23975 24074: gap of 100 bp
* 24075 24745: contig of 671 bp in length
* 24746 24845: gap of 100 bp
* 24846 25522: contig of 677 bp in length
* 25523 25622: gap of 100 bp
* 25623 26294: contig of 672 bp in length
* 26295 26394: gap of 100 bp
* 26395 27061: contig of 667 bp in length
* 27062 27161: gap of 100 bp
* 27162 27827: contig of 666 bp in length
* 27828 27927: gap of 100 bp
* 27928 28611: contig of 684 bp in length
* 28612 28711: gap of 100 bp
* 28712 29395: contig of 684 bp in length

```

```

* 29396 29495: gap of 100 bp
* 29496 30166: contig of 671 bp in length
* 30167 30266: gap of 100 bp
* 30267 30946: contig of 680 bp in length
* 30947 31046: gap of 100 bp
* 31047 31700: contig of 654 bp in length
* 31701 31800: gap of 100 bp
* 31801 32455: contig of 655 bp in length
* 32456 32555: gap of 100 bp
* 32556 33224: contig of 669 bp in length
* 33225 33324: gap of 100 bp
* 33325 34012: contig of 688 bp in length
* 34013 34112: gap of 100 bp
* 34113 34756: contig of 644 bp in length
* 34757 34856: gap of 100 bp
* 34857 35514: contig of 658 bp in length
* 35515 35614: gap of 100 bp
* 35615 36297: contig of 683 bp in length
* 36298 36397: gap of 100 bp
* 36398 37066: contig of 669 bp in length
* 37067 37166: gap of 100 bp
* 37167 37848: contig of 682 bp in length
* 37849 37948: gap of 100 bp
* 37949 38606: contig of 658 bp in length
* 38607 38706: gap of 100 bp
* 38707 39385: contig of 679 bp in length
* 39386 39485: gap of 100 bp
* 39486 40178: contig of 693 bp in length
* 40179 40278: gap of 100 bp
* 40279 40935: contig of 657 bp in length
* 40936 41035: gap of 100 bp
* 41036 41718: contig of 683 bp in length
* 41719 41818: gap of 100 bp
* 41819 42492: contig of 674 bp in length
* 42493 42592: gap of 100 bp
* 42593 43275: contig of 683 bp in length
* 43276 43375: gap of 100 bp
* 43376 44038: contig of 663 bp in length
* 44039 44138: gap of 100 bp
* 44139 44821: contig of 683 bp in length
* 44822 44921: gap of 100 bp
* 44922 45602: contig of 681 bp in length
* 45603 45702: gap of 100 bp
* 45703 46382: contig of 680 bp in length
* 46383 46482: gap of 100 bp
* 46483 47153: contig of 671 bp in length
* 47154 47253: gap of 100 bp
* 47254 47933: contig of 680 bp in length
* 47934 48033: gap of 100 bp
* 48034 48705: contig of 672 bp in length
* 48706 48805: gap of 100 bp
* 48805 49481: contig of 676 bp in length
* 49482 49581: gap of 100 bp
* 49582 50269: contig of 688 bp in length.

FEATURES
    Location/Qualifiers
        1..50269
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /clone="RP24-483K3"
            /clone_lib="RPCI-24 Male Mouse BAC"
BASE COUNT 12034 a 9403 c 9549 g 12707 t 6576 others
ORIGIN

Alignment Scores:
Pred. No.: 9.83e-06 Length: 50269
Score: 16.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.08% Indels: 0
DB: 2 Gaps: 0

US-09-966-880A-8 (1-198) x AC119975 (1-50269)

```

```

QY      26 GluThrTyrLeuCystrValVallYsArgAspSerAlaThrSer 41
      |||||||
Db      35893 GAGACCTACCTCTGCTACGTGTGAGAGAGAGAGATAGTCCACCTCC 35846

RESULT 53
AC068309/c      67244 bp      DNA      linear      HTG 16-OCT-2001
DEFINITION      Mus musculus chromosome 13 clone RP23-298J20, *** SEQUENCING IN
PROGRESS      *** 22 unordered pieces.
ACCESSION      AC068309
VERSION      HTG; HTGS_PHASE1.
KEYWORDS      Mus musculus.
SOURCE      Mus musculus.
ORGANISM      Mus musculus
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.
1 (bases 1 to 67244)
Metzker,M.L., Lewis,L.R., Hume,J., Edwards,C., Harris,C.,
Dederich,D., Thomas,S., Okwuonu,G., Carlock,C., Garner,T.,
Addison,S., Pace,A., Williams,G., Bonnin,D., Brooks,A., Brown,J.,
Buhay,C., Bunac,C., Burkett,C., Chacko,J., Chen,G., Chen,Z.,
Cox,C., Davis,C., Delgado,O., Ding,Y., Dugan-Rocha,S.,
Fernandez,C., Ferraguto,D., Forcum-Tansey,J., Gill,R.,
Gorell,J.H., Gunaratne,P., Haller,G., Hernandez,J., Hognes,M.,
Hosak,H., Hou,X., Huber,J., Jackson,L., Jia,Y., Kelly,J., Kelly,S.,
Kovar,C., Liu,J., Liu,W., Loulseghe,H., Lozado,R.J., Martin,R.,
Massey,E., McLeod,M.P., Mei,G., Moore,S., Morgan,M., Morris,S.,
Neal,D., Nelson,A., Nguyen,R., Nguyen,N., Ogih,M., Parish,B.,
Perez,L., Reiter,D., Say,J., Shen,H., Vasquez,L., Watlington,S.,
Williamson,A., Wrensford,G., Zhou,X., Bouck,J., Hodgson,A.,
Muzny,D.M., Rives,M., Scherer,S., Sodergren,E., Weinstock,G.,
Worley,K. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 67244)
Worley,K.C.
Direct Submission
Submitted (02-MAY-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Oct 14, 2001 this sequence version replaced gi:11094637.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: MADZ
Center clone name: RP23-298J20
----- Summary Statistics
Sequencing vector: M13: L08821
Chemistry: Dye-primer Bodipy: 100% of reads
Assembly program: Phrap: version 0.990329
Consensus quality: 69254 bases at least Q40
Consensus quality: 93836 bases at least Q30
Consensus quality: 103881 bases at least Q20
Estimated insert size: 89529: sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-gel estimation
Quality coverage: 1x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 22 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 4843: contig of 4843 bp in length
* 4844 4943: gap of unknown length
* 4944 9846: contig of 4903 bp in length

```

```

*      9847 9946: gap of unknown length
*      9947 15031: contig of 5085 bp in length
*      15032 15131: gap of unknown length
*      15132 18316: contig of 3185 bp in length
*      18317 18416: gap of unknown length
*      18417 21312: contig of 2896 bp in length
*      21313 21412: gap of unknown length
*      21413 25595: contig of 4183 bp in length
*      25596 25695: gap of unknown length
*      25696 28983: contig of 3288 bp in length
*      28984 31371: contig of 2288 bp in length
*      31372 31471: gap of unknown length
*      31472 34502: contig of 3031 bp in length
*      34503 34603: gap of unknown length
*      34603 37164: contig of 2562 bp in length
*      37165 37264: gap of unknown length
*      37265 40058: contig of 2794 bp in length
*      40059 40158: gap of unknown length
*      40159 42440: contig of 2282 bp in length
*      42441 42540: gap of unknown length
*      42541 44901: contig of 2361 bp in length
*      44902 45002: gap of unknown length
*      45003 47782: contig of 2781 bp in length
*      47783 47882: gap of unknown length
*      47883 50069: contig of 2187 bp in length
*      50070 50169: gap of unknown length
*      50170 52502: contig of 2333 bp in length
*      52503 52603: gap of unknown length
*      52603 55087: contig of 2485 bp in length
*      55088 55187: gap of unknown length
*      55188 57301: contig of 2114 bp in length
*      57302 57401: gap of unknown length
*      57402 59994: contig of 2593 bp in length
*      59995 60094: gap of unknown length
*      60095 62360: contig of 2266 bp in length
*      62361 62460: gap of unknown length
*      62461 65125: contig of 2665 bp in length
*      65126 65225: gap of unknown length
*      65226 67244: contig of 2019 bp in length.
FEATURES
      Location/Qualifiers
          1..67244
              /organism="Mus musculus"
              /db_xref="taxon:10090"
              /chromosome="13"
              /clone="RP23-298J20"
BASE COUNT      17937 a 13995 c 14816 g 18342 t 2154 others
ORIGIN
Alignment Scores:
Pred. No.:      30.1      Length:      67244
Score:          10.00     Matches:    10
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      5.05%      Indels:    0
DB:              2      Gaps:    0
US-09-966-880A-8 (1-198) x AC068309 (1-67244)
QY      169 SerValArgLeuSerArgGlnLeuArgArg 178
      |||||||
Db      12769 TCTGTGAGGCTGAGCAGGACGCTTCGTAGA 12740
RESULT 54
DMSPALTR
LOCUS      D.melanogaster spalt-related gene.
DEFINITION      Y07653
ACCESSION      Y07653
VERSION      Y07653.1 GI:1684680
KEYWORDS      spalt gene; zinc finger.
SOURCE      Drosophila melanogaster.
ORGANISM      Eukaryota: Metazoa: Arthropoda: Hexapoda: Insecta: Pterygota:

```





```

GQSGKPPDPYDEASSKRKARDCVVALMDHFGFNEFFSLAGHDRGSAFTAFRTAMDHPAR
VKLAILDVGLLEALEHCNARFATQWWHFEFFAQHDKPEQAI SADPLRYWYGGTEDLM
GKNVADFLAVQDHSTIQMLGDYRAGVRIDHLDHREDRNSGKLOCPLRLVLSLRD
DLEDLYNDVLAIWPAELPVSGLGIDSGHHMAEAPLEVARELKSFFLA"
complement(1637. .2269)
/gene="AGR_pat_729"
CDS
/gene="AGR_pat_729"
complement(1637. .2269)
/translation="putative tetra family transcriptional
regulator"
/codon_start=1
/transl_table=11
/product="AGR_pat_729p"
/protein_id="AAK90871.1"
/db_xref="GI:15162312"
/translation="MTCDSHNSWSALHVPKERTQPLSEAPLRADARNRKLIE
VASAFESQGAASLEEDTARAGVGIGTLYRHFPTRHLSVESYRRELELATAARDL
IAEEPAPAI EEMNRFYVMAARKGMSLKLFSTNSITFAEGSKLLHTAFEGLLD
NAVADAVKGIDIEADVLNALGIYSIPEGPWRDRAHVRVLVMDGLRR"
2444. .3433
/gene="AGR_pat_731"
CDS
2444. .3433
/gene="AGR_pat_731"
/translation="HYPOTHETICAL 35.0 KD PROTEIN IN INTF-EAEH
INTERGENIC REGION"
/codon_start=1
/transl_table=11
/product="AGR_pat_731p"
/protein_id="AAK90872.1"
/db_xref="GI:15162313"
/translation="MTCDSADIRHNGAPRWLPNADIPVAVPQKAFSTDTACDIIS
FAIANIAGLVGCTIVIEIRGSSRPLGAHMAVQDRCYGYSGGCTEAAVAEAL
QAMSGHDFVLLGCGFPFDVLPKGGITVAIHLKDVDPCLCSVNLVDRGSATL
QCYGLELSGSCGARTGPDGDFLPSYRKPRIEGRSLEVEYVAVAKAGYEV
IFEDHTAGDSFDSLIDATVAILLHDLDEMPILGALVATPFYIGALGSSRIHEK
RYERURQIGHSEAIARIPGILFGRADAQSLAFSVVADIAASRSCVMHS"
3561. .4094
/gene="AGR_pat_732"
CDS
3561. .4094
/gene="AGR_pat_732"
/translation="Oxidoreductase, iron-sulfur subunit - Deinococcus
radiodurans (strain R1)"
/codon_start=1
/transl_table=11
/product="AGR_pat_732p"
/protein_id="AAK90873.1"
/db_xref="GI:15162314"
/translation="MORTINGAVEADINTSLDLNRNVLGFTGKKGCDQACGA
CTVLVDGERINSLALAVQYGRSITTVGLAANGNLHPLOKAFIEHGDGQCYCTPG
QLCSAIGMAGEIERNIPSVVTADLSNSITVDDTEIRERMSGNLCRCGAYNGIVEATS
DTLVQQTAVPSEERARA"
4091. .5080
/gene="AGR_pat_735"
CDS
4091. .5080
/gene="AGR_pat_735"
/translation="Oxidoreductase - Deinococcus radiodurans (strain
R1)"
/codon_start=1
/transl_table=11
/product="AGR_pat_735p"
/protein_id="AAK90874.1"
/db_xref="GI:15162315"
/translation="MNFISYHQPSDTASIALKQAGPTAKYLAGGTTNLVDMRETVER
PETLVDTHTLSSITERTDGLLVAGVKNALAAHDAVRAQFPLLSRALAGASAQI
RNATVGNILQTRCFYFVNAARCNKRONPSGCDALGEGNRYHATILGASECVATH
PSDCVALVALDASVHLQGPAGTRSLPLRLHRLPGHTPEIETELDTDDLLTAVEIPS
LPMFGATYRKVRDASTAFALVSAALVDVDDAGNVCYKRLALGGVVAHPKRAKAE
SLRKQQAASERAAEALADLAIGLQDNFAKTELAKRTIVAVLEELKGESA"
5077. .7380
/gene="AGR_pat_739"
CDS
5077. .7380
/gene="AGR_pat_739"
/translation="probable oxidoreductase (EC 1.-.-.-) DRA0237

```

```

(similarity) - Deinococcus radiodurans (strain R1)"
/codon_start=1
/transl_table=11
/product="AGR_pat_739p"
/protein_id="AAK90875.1"
/db_xref="GI:15162316"
/translation="MSTDOSOTLEAGKLGWQPAVTSDDLRLKRIHAGLGGISRIEGPM
KVQCTRFPAEEFYDNIYSALAASTVARGVRIDLVNSAAEAHGVILVMTKNAPRM
KAPLSMSSPTAGASNLPMQSDIEHNGOPIALVIAETOEDHAASLVTKAYTELL
PAVTSFDDAKKSPHOLDNLGQPPFIDIGDAETALANAEXKVDLVYRPRHNAIEL
HAATVVMKDDLRVDSQLDLITGQLADIFGLDVSXVHVTSYVGGFGGKGFMDH
QILACAAARLAGRPVRLMREGVFRIGGRTVTEQORVALCAKADGTLDALIHTGTA
MTLHNSCEQFTFPAHLYAAKTFHIGQDVADMDMLANTFMRAFGESVGTFALECALD
ELAEKGLDPIELRRTEPEKPTTGKPFSSRYLLEAYEKGAEOFGHKSQTPROKR
EGEWLIGMCATATYPYHFRPGGAVRIKLTVDGRVTVSTAVHDMGMDGTATQMOHILAA
RLGLPLDHVTFEYGDLSKPRVITAGSTQSTASIGGAVIAATEVIVEELJLISGNDSP
AGLSLLEVARDDGLSHISDNLSFESYSLRRAGREELVCEAPAPAPAEAPFMSHS
YGAQFCVHVSAARTGETRVSRFLGSDAGQILNSKMATSQFKGII MGICGLALTETN
FDBRTGRIVNASLADYHVPQMDVPAIEILYTNKPDQAPANGANGICEIGITGVGAAY
ANAVYNATGVVRDLPITLLOKLTGGLD"
complement(7724. .8416)
/gene="AGR_pat_742"
CDS
complement(7724. .8416)
/gene="AGR_pat_742"
/translation="conserved hypothetical protein - Deinococcus
radiodurans (strain R1)"
/codon_start=1
/transl_table=11
/product="AGR_pat_742p"
/protein_id="AAK90876.1"
/db_xref="GI:15162317"
/translation="MRNMNVGMMKLTETRSPQDKAEGNRPPVTVAIIILAAGKASR
MGDGAHKLIAEDGKPLVRSFAIALTADAASVIVVTGHRRSFIEAALLGLGVTTN
NPDYASGASLSLAGFAPACADGVLVWLADMPVTIGDLDLIIISAPRMSRGLV
VYRSGRGRGNPVLIPRLSLNGAVMRLEGDVGARQTIETSGRLIIVETIGDSALHDVDP
EAVVAAGGIIRT"
complement(8501. .9256)
/gene="AGR_pat_743"
CDS
complement(8501. .9256)
/gene="AGR_pat_743"
/translation="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_pat_743p"
/protein_id="AAK90877.1"
/db_xref="GI:15162318"
/translation="MIKRTIAVLTIVATTTWCATSSNAREGIVRVAVYKGPASCEDC
SETMRKISIMALGPEYSVDYVGAEEKIDISDETLSKFDIYVOPGGODIPAAFPESLGEQ
RVEAIRFVAGGGFGSCMGAYLAGSGIGLIDDELESEVGRFGFPVSSIISTAVST
SWNGEKOTVFQDQPYMKPEKDIRGFKKIATYQNGDIAAARYHFGNGIVVLGTGHPHA
GGSWFHEAGIPIDKMPDHRVLRVMDERSEPR"
complement(9366. .10823)
/gene="AGR_pat_745"
CDS
complement(9366. .10823)
/gene="AGR_pat_745"
/translation="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_pat_745p"
/protein_id="AAK90878.1"
/db_xref="GI:15162319"
/translation="MPSRKLKLSLTPFAALASPLVQOOTVAATIVSISMGQAVAG
SPVRMNPALLGDMRRSLAEAFSEAKGDSKRRSVAVLAAITATQAVQSILETAI
AANDASIRATKALSHAVGNLOTYSLSAKCADVTRALORFNAWAAYSSRFLVAK
HKASAAATASDQVALRRKRVADLSYRMOQLEDVADNAALRGEVIRMRREAIITOSTIP
ANAYDQRLSFTLTVLGSPFAALSITRTYTPYSVADPEVRSIRITTDYDGYDGY
YQGVADYSYSEPVISSPVSTETRTYQNTYETIYSVTNETTNIYQTMPEDELSDV
YQVVPQGVNFVEDISANKSGSEPVDEFSTYGLPASEAASGTMPEVLPSPNSDQA
PDVPPDODEPPPAIPSPLEHNSDNGAGTFDDTAGSVSPDHDDKIPWAPEDRISDIHST
PDTLPSSENDEPRAGRVQEDSEEPVWKATP"
2307 a 3145 c 3105 g 2388 t

```

BASE COUNT  
ORIGIN

## Alignment Scores:

Pred. No.: 72.4 Length: 10945  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 4.55% Indels: 0  
 DB: 1 Gaps: 0

US-09-966-880A-8 (1-198) x AE007918 (1-10945)

Qy 125 GlyLeuAtgArgLeuHisArgAlaGly 133

Db 3680 GGCCTGCGCGCGCTGCACCGTGTGTGT 3706

## RESULT 56

AE008970

LOCUS

DEFINITION Agrobacterium tumefaciens str. C58 AT plasmid section 45 of 49 of the complete sequence.

ACCESSION AE008970 AE008687

VERSION AE008970.1 GI:17743952

KEYWORDS

SOURCE

ORGANISM

REFERENCE Agrobacterium tumefaciens str. C58 (U. Washington).

AUTHORS Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group; Rhizobiaceae; Rhizobium.

1 (bases 1 to 11013)

Wood, D.W., Setubal, J.C., Kaul, R., Monks, D., Chen, L., Wood, G.E., Zhou, Y., Woo, L., Kitajima, J.P., Okura, V.K., Almeida Jr., N.F., Zhou, Y., Bovee Sr., D., Chapman, P., Clendenning, J., Deatherage, G., Gillet, W., Grant, C., Guenther, D., Kutayavin, T., Levy, R., Li, M., McClelland, E., Palmieri, A., Raymond, C., Rouse, G., Saenphimmachak, C., Wu, Z., Gordon, D., Eisen, J.A., Paulsen, I., Karp, P., Romero, P., Zhang, S., Yoo, H., Tao, Y., Biddle, P., Jung, M., Krespan, W., Perry, M., Gordon-Kamm, B., Liao, L., Kim, S., Hendrick, C., Zhao, Z., Dolan, M., Tingey, S.V., Tomb, J., Gordon, M.P., Olson, M.V., and Nester, E.W.

TITLE The genome of the natural genetic engineer Agrobacterium

JOURNAL tumefaciens C58

MEDLINE Science 294 (5550), 2317-2323 (2001)

PUBMED 21608550

11743193

2 (bases 1 to 11013)

Wood, D.W., Setubal, J.C., Kaul, R., Monks, D., Chen, L., Wood, G.E., Zhou, Y., Woo, L., Kitajima, J.P., Okura, V.K., Almeida Jr., N.F., Zhou, Y., Bovee Sr., D., Chapman, P., Clendenning, J., Deatherage, G., Gillet, W., Grant, C., Guenther, D., Kutayavin, T., Levy, R., Li, M., McClelland, E., Palmieri, A., Raymond, C., Rouse, G., Saenphimmachak, C., Wu, Z., Gordon, D., Eisen, J.A., Paulsen, I., Karp, P., Romero, P., Zhang, S., Yoo, H., Tao, Y., Biddle, P., Jung, M., Krespan, W., Perry, M., Gordon-Kamm, B., Liao, L., Kim, S., Hendrick, C., Zhao, Z., Dolan, M., Tingey, S.V., Tomb, J., Gordon, M.P., Olson, M.V., and Nester, E.W.

Direct Submission

JOURNAL Submitted (27-SEP-2001) Department of Microbiology, University of

Washington, 1959 NE Pacific Ave, Box 357242, Seattle, WA

98195-7242, USA

## FEATURES

source

1..11013 Location/Qualifiers

/organism="Agrobacterium tumefaciens str. C58 (U.

Washington)"

/strain="C58"

/db\_xref="taxon:180835"

/plasmid="At"

538..1398

/gene="Atu5495"

538..1398

/gene="Atu5495"

/note="identified by sequence similarity; putative; ORF

located using Blastx/Glimmer"

/codon\_start=1

/transl\_table=11

/product="hydrolase"

gene

CDS

/protein\_id="AAL46181.1"  
 /db\_xref="GI:17743953"  
 /translation="MFENFRLEMKVLPETKLRVRHGGSPPLLLLHGHPRTHTTWYKV  
 APIIAEHFTVICPDLPFGQSGKDDPDYDEASSKRAKARDCAVLMQDHFGEFSLAG  
 HDRGSYTAFTAMDHPARVAKAILLDGVPILEALEHCNAREATQWHHFFAQHDKPE  
 QAISADPLRWYGGTDLGMKNYADFLVAVQDDHSTIQGMLGDYRAGVRIHLDHREDR  
 NSRKLCPLRLVLSLRDLDLEDLYNDVLAIQPWAELPVSGLGIDSGHHMAEAPLEY  
 ARELKSFFLA"  
 complement(1636..2220)  
 /gene="Atu5496"  
 complement(1636..2220)  
 /gene="Atu5496"  
 /note="identified by sequence similarity; putative; ORF  
 located using Blastx/Glimmer"  
 /codon\_start=1  
 /transl\_table=11  
 /product="transcriptional regulator, TetR family"  
 /protein\_id="AAL46182.1"  
 /db\_xref="GI:17743954"  
 /translation="MPKERTQPPLEAPLRADARRNREKLIEVASLAFSEQGAASLE  
 DIARRAGVGIGTLVRFPTREHLVSVYRRELELLATAARDLIAEPEAEATEEWMNR  
 FVNYMAAKGMANSIKLLFTSNSTLFAESKLLHTAFEGLLDNVAKADVKGDIIEAAD  
 VLNALFGIYSIPEGPEDRAHRIVRLVMDGLRR"  
 2443..3432  
 /gene="Atu5497"  
 2443..3432  
 /gene="Atu5497"  
 /note="identified by sequence similarity; putative; ORF  
 located using Blastx/Glimmer"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AAL46183.1"  
 /db\_xref="GI:17743955"  
 /translation="MTQSADIIRHNGGAPRLPRNADIPYAVVPKAFSTDTACDILS  
 FANIKDRGLQVGLCTLVEIRGGSRPLGAHMAVQDGRYCGYVSGGCTEAAVAEAL  
 QAMKSGHDFVLLGEGSPFFDIVPCGGITVAIHLKDDVPLCVSNMLVDGRRSATL  
 CYSGLLELSLGSFGARTGWDGIDFLTYSRPKPRVIFSGRLEVEVAVNAKAGGYEV  
 IFFDTAGDSFSSLDIADTAVALLHDLLEMPILGAALVATPFYIGALGSSRTHKEK  
 RVERLROIGHSEEAIRAKIPIGLFGARDAQSLAFSVVADIAASRQSCVMHS"  
 3560..4093  
 /gene="Atu5498"  
 3560..4093  
 /gene="Atu5498"  
 /note="identified by sequence similarity; putative; ORF  
 located using Blastx/Glimmer"  
 /codon\_start=1  
 /transl\_table=11  
 /product="oxidoreductase with iron-sulfur subunit"  
 /protein\_id="AAL46184.1"  
 /db\_xref="GI:17743956"  
 /translation="MQFTINGFAVEADIFTLLDLMRNYLGFCTKKGCDGACGA  
 CTVLVDGRINSCLALAVOYQGRSITTVTEGLAANGNLHPLOKAFIEHGDGFCGCTPG  
 QLCSAIGHAGEIERNIPSVVTADLSSNITVDDTEIRERMSGNLRCGAYNGIVEAIS  
 DTLVQQTAVPSEERARA"  
 4090..5079  
 /gene="Atu5499"  
 4090..5079  
 /gene="Atu5499"  
 /note="identified by sequence similarity; putative; ORF  
 located using Blastx/Glimmer"  
 /codon\_start=1  
 /transl\_table=11  
 /product="oxidoreductase"  
 /protein\_id="AAL46185.1"  
 /db\_xref="GI:17743957"  
 /translation="MNIIFYHQPDSATAIALKQAGPTAKYLAGTNLVDLMRETVR  
 PETLVDVTHLSIIERRTDGSLLVAGVKNALAAADAVRAQVPLLSRLAAGASAOI  
 RNVAIVGGNIIQRTCRVFDYNAARKRQPNQSGDALEGFNRYHAIIICASGCVATH  
 PSDMCVALVALDASVHLOGPACTRSILPLDLHRLPCHTPEIETELDLDLITAVEIPS  
 LPFAGRSYRKVRDRASAFALVSVAALDVDDAGNVKDVRLALGVAHKPWRALKAE  
 SRLKQQAESFRAAEAEALADAIGLQDNAPFKIELAKRTIVAVLEELKGESA"  
 5076..7379

gene

```
CDS
/gene="Atu5500"
5076..7379
/feature="Atu5500"
/note="Identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="dehydrogenase"
/protein_id="AAL46186.1"
/db_xref="GI:17743958"
/translation="MSTDSQTLPEAGKLGKRMQPAVTSDDLRLKHGALQGQISRIEGPM
KYQGTREAFPPYDNIISYAALFSTVARGVRIDLVNAAEAAPGVILVMTHTKNAPRM
KAPSLMSPATAAGASNLPMQSDIHWNGOPIALVLAETQDAADAAASLTKAYELL
PAVTSFDDAKSPROLDNLGPPFDIDGDAETLANAEVKVDLVRTPRNHNAIEL
HAATVYWDDELVRHDAEQDLDTLTCGLADIFGLDYSKVHVTSPPVGGGFGKCFWDH
QLIACAARAGLRGPRVIMLSREGVFRIGGRVTVEQRVALGAKADGTLDLALITGAA
MTLHNSCEPEQTFPAHRIYAAKTFHGQDVADMDMLANTFMRAPESVGTFALECALD
ELAKLIGDPIELRRIDPEKDTPTGKPFSSRIIEAYEAKGAEQGWKRSQTPQKR
EGEWLIGMCATATPYHRFGPGAVRIKLTVDGRVTSTAVHDMGMGTATQMOHLAA
RLGLPLDHTVTFEYGSKLPRGVIAGSQTQISIGGAVIAATEVLVEELIKLGNDSPL
AGLSLLEVEARDGGLSHLSDNLSFESYQSLRRAGREELVCEAPAPAEAFSMHS
YGAQCEVHVSARTGRTVRFGLGSDAGIILNSKMATSFQKGLIMIGIGLALTEETN
FDQRTGRIVNASLDYHVPVQMDVPAIEILYTNKPDPOAPMGARGIGEIGITGVGAAY
ANAVYNATGVRVRDLPITLKLITGLD"
complement(7723..8385)
/feature="Atu5501"
complement(7723..8385)
/feature="Atu5501"
/note="Identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAL46187.1"
/db_xref="GI:17743959"
/translation="MKITRPSQDKAEGRNRPVTVAAIILAAKASRMGDGGAHKLL
AFEDKPIVRSAAELADRAASVIVTGHRRSEAEALLGIVGVTNNPDVYASGIAS
SLIAGFSAPACDADGVLMADMEFVTIGDLDTLISAFRMASSRAIVRAVSRGKRG
PVILPRSLNGAVMRLEGDVGARQIETSLGRIIDIVEIGDSAHLDVDTPEAVVAAGIL
RT"
complement(8500..9255)
/feature="Atu5502"
complement(8500..9255)
/feature="Atu5502"
/note="Identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAL46188.1"
/db_xref="GI:17743960"
/translation="MIKTIATLTVIATTWCATSSNAREGIVRVAVYRGPAQCEDC
SETMRKSLMAGPEYSDYVGAEEKIDIDSETLSEFDIYVOPGGQDIPAAVESLGEQ
RVEALRRFVEAGGFFGCMGAYLASGSGIGLIDDELESEVGRGPPVSSIIISTAVST
SWNGEKQTVFFQDGPYMSPEKDKIRGFKKIATYQNGIDIAAARYHFGNGIVLVTPHPEA
GGSWFHEAGIPIDKMPDHRVLRVSNVDEFRSEPR"
complement(9365..10822)
/feature="Atu5503"
complement(9365..10822)
/feature="Atu5503"
/note="Identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAL46189.1"
/db_xref="GI:17743961"
/translation="MPRSKLVSLTLPFAALAASPLVQQTVAATAIVSLSMQAVAG
SPVRNAPALLGDMRSRLAEIAFSESEAKGDSKRSRVAVLAALTEATQAVOSLETAI
AAHDAASRISRATKALSHAVGNLQTRYSLSSAKCDVATRALQRFNAWAAYESSRFLAK
```

Alignment Scores:  
Pred. No.: 72.8 Length: 11013

```
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.55% Indels: 0
DB: 1 Gaps: 0

US-09-966-880A-8 (1-198) x AE008970 (1-11013)

QY 125 GlyLeuAAGAgLeuHisArgAlaGly 133
Db 3679 GGCATTGGCGGCGCTGCACCGTCTCGGT 3705

RESULT 57
AC100984 59915 bp DNA linear HTG 23-NOV-2001
LOCUS Mus musculus clone RP23-76J21, LOW-PASS SEQUENCE SAMPLING.
DEFINITION AC100984
ACCESSION AC100984
VERSION AC100984.1 GI:17059758
HTG: HTGS_PHASE0.
KEYWORDS Mus musculus.
SOURCE Mus musculus.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 59915)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Mus musculus, clone RP23-76J21
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 59915)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., All,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepei,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyne,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Jones,B., Heatford,A., Horton,L., Hulme,W., Illiev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRoque,K.,
Lamazares,R., Landers,T., Lehoczyk,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
Meneus,L., Mhova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,N., Schupbach,K.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trifilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L14683
Center clone name: 76_J_21
-----
* NOTE: This record contains 74 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for:
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
```

\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 695: contig of 695 bp in length  
\* 696 795: gap of 100 bp  
\* 796 1494: contig of 699 bp in length  
\* 1495 1594: gap of 100 bp  
\* 1595 2301: contig of 707 bp in length  
\* 2302 2401: gap of 100 bp  
\* 2402 3107: contig of 706 bp in length  
\* 3108 3207: gap of 100 bp  
\* 3208 3905: contig of 698 bp in length  
\* 3906 4005: gap of 100 bp  
\* 4006 4712: contig of 707 bp in length  
\* 4713 4812: gap of 100 bp  
\* 4813 5519: contig of 707 bp in length  
\* 5520 5619: gap of 100 bp  
\* 5620 6336: contig of 717 bp in length  
\* 6337 6436: gap of 100 bp  
\* 6437 7155: contig of 719 bp in length  
\* 7156 7255: gap of 100 bp  
\* 7256 7980: contig of 725 bp in length  
\* 7981 8080: gap of 100 bp  
\* 8081 8798: contig of 718 bp in length  
\* 8799 8998: gap of 100 bp  
\* 8999 9616: contig of 718 bp in length  
\* 9617 9716: gap of 100 bp  
\* 9717 10436: contig of 720 bp in length  
\* 10437 10536: gap of 100 bp  
\* 10537 11266: contig of 730 bp in length  
\* 11267 11366: gap of 100 bp  
\* 11367 12073: contig of 707 bp in length  
\* 12074 12173: gap of 100 bp  
\* 12174 12874: contig of 701 bp in length  
\* 12875 12974: gap of 100 bp  
\* 12975 13673: contig of 699 bp in length  
\* 13674 13773: gap of 100 bp  
\* 13774 14447: contig of 674 bp in length  
\* 14448 14547: gap of 100 bp  
\* 14548 15251: contig of 704 bp in length  
\* 15252 15351: gap of 100 bp  
\* 15352 16052: contig of 701 bp in length  
\* 16053 16152: gap of 100 bp  
\* 16153 16873: contig of 721 bp in length  
\* 16874 16973: gap of 100 bp  
\* 16974 17694: contig of 721 bp in length  
\* 17695 17794: gap of 100 bp  
\* 17795 18516: contig of 722 bp in length  
\* 18517 18616: gap of 100 bp  
\* 18617 19334: contig of 718 bp in length  
\* 19335 19434: gap of 100 bp  
\* 19435 20155: contig of 721 bp in length  
\* 20156 20255: gap of 100 bp  
\* 20256 20961: contig of 706 bp in length  
\* 20962 21061: gap of 100 bp  
\* 21062 21764: contig of 703 bp in length  
\* 21765 21864: gap of 100 bp  
\* 21865 22563: contig of 699 bp in length  
\* 22564 22663: gap of 100 bp  
\* 22664 23383: contig of 720 bp in length  
\* 23384 23483: gap of 100 bp  
\* 23484 24206: contig of 723 bp in length  
\* 24207 24306: gap of 100 bp  
\* 24307 25024: contig of 718 bp in length  
\* 25025 25124: gap of 100 bp  
\* 25125 25820: contig of 696 bp in length  
\* 25821 25920: gap of 100 bp  
\* 25921 26654: contig of 734 bp in length  
\* 26655 26754: gap of 100 bp  
\* 26755 27450: contig of 696 bp in length  
\* 27451 27550: gap of 100 bp  
\* 27551 28255: contig of 705 bp in length  
\* 28256 28355: gap of 100 bp

28356 29049: contig of 694 bp in length  
29050 29149: gap of 100 bp  
29150 29853: contig of 704 bp in length  
29854 29953: gap of 100 bp  
29954 30670: contig of 717 bp in length  
30671 30770: gap of 100 bp  
30771 31474: contig of 704 bp in length  
31475 31574: gap of 100 bp  
31575 32287: contig of 713 bp in length  
32288 32387: gap of 100 bp  
32388 33115: contig of 728 bp in length  
33116 33215: gap of 100 bp  
33216 33925: contig of 710 bp in length  
33926 34025: gap of 100 bp  
34026 34748: contig of 723 bp in length  
34749 34848: gap of 100 bp  
34849 35542: contig of 694 bp in length  
35543 35642: gap of 100 bp  
35643 36326: contig of 684 bp in length  
36327 36426: gap of 100 bp  
36427 37132: contig of 706 bp in length  
37133 37232: gap of 100 bp  
37233 37955: contig of 723 bp in length  
37956 38055: gap of 100 bp  
38056 38768: contig of 713 bp in length  
38769 38868: gap of 100 bp  
38869 39572: contig of 704 bp in length  
39573 39672: gap of 100 bp  
39673 40401: contig of 729 bp in length  
40402 40501: gap of 100 bp  
40502 41211: contig of 710 bp in length  
41212 41311: gap of 100 bp  
41312 42044: contig of 733 bp in length  
42045 42144: gap of 100 bp  
42145 42846: contig of 702 bp in length  
42847 42946: gap of 100 bp  
42947 43655: contig of 709 bp in length  
43656 43755: gap of 100 bp  
43756 44459: contig of 704 bp in length  
44460 44559: gap of 100 bp  
44560 45262: contig of 703 bp in length  
45263 45362: gap of 100 bp  
45363 46066: contig of 704 bp in length  
46067 46166: gap of 100 bp  
46167 46887: contig of 721 bp in length  
46888 46987: gap of 100 bp  
46988 47708: contig of 721 bp in length  
47709 47808: gap of 100 bp  
47809 48509: contig of 701 bp in length  
48510 48609: gap of 100 bp  
48610 49333: contig of 724 bp in length  
49334 49433: gap of 100 bp  
49434 50152: contig of 719 bp in length  
50153 50252: gap of 100 bp  
50253 51065: contig of 813 bp in length  
51066 51165: gap of 100 bp  
51166 51842: contig of 677 bp in length  
51843 51942: gap of 100 bp  
51943 52652: contig of 710 bp in length  
52653 52752: gap of 100 bp  
52753 53461: contig of 709 bp in length  
53462 53561: gap of 100 bp  
53562 54264: contig of 703 bp in length  
54265 54364: gap of 100 bp  
54365 55054: contig of 690 bp in length  
55055 55154: gap of 100 bp  
55155 55882: contig of 728 bp in length  
55883 55982: gap of 100 bp

Alignment Scores:

Pred. No.: 315 Length: 59915  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0

```

Best Local Similarity: 100.00%      Mismatches: 0
Query Match:          4.55%        Indels: 0
DB:                   2            Gaps: 0

US-09-966-880A-8 (1-198) x AC100984 (1-59915)
Qy  58 GluLeuLeuPheLeuArgTyrlSer 66
    |||||
Db  51017 GAATTGTTATTTTGAGATACATCTCC 51043

RESULT 58
AC014357/c
LOCUS AC014357 103746 bp DNA linear HTG 16-NOV-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.
ACCESSION AC014357
VERSION AC014357.1 GI:6436978
KEYWORDS HTG; HTGS_PHASE2.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 103746)
Adams,M. and Venter,J.C.
Direct Submission
Rockville, MD, USA
This sequence was identified as CDM:10212654 by the submitter.
For further information on this sequence e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES             Location/Qualifiers
     source            1..103746
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT  29847 a 21798 c 21974 g 30127 t
ORIGIN
Alignment Scores:
Pred. No.:          506      Length:      103746
Score:              9.00      Matches:      9
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match:        4.55%     Indels:      0
DB:                 2         Gaps:        0

US-09-966-880A-8 (1-198) x AC014357 (1-103746)
Qy  117 GluAspArgLysAlaGluProGluGly 125
    |||||
Db  55591 GAGACACAGAGCGAGACCCAGAGCGT 55565

RESULT 59
AC087457
LOCUS AC087457 141899 bp DNA linear PRI 30-NOV-2001
DEFINITION Homo sapiens chromosome 15, clone RP11-814P5, complete sequence.
ACCESSION AC087457
VERSION AC087457.5 GI:17155075
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 141899)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Unpublished
JOURNAL Homo sapiens chromosome 15, clone RP11-814P5
2 (bases 1 to 141899)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B., Brown,A.,
Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Furo,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., Larocque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meidrim,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nanyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Riley,R., Rise,C., Rogov,P.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigillo,J., Vassiliev,H.,
Viel,R., Vo.A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (08-SEP-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 141899)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Furo,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., Larocque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meidrim,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nanyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigillo,J., Vassiliev,H.,
Viel,R., Vo.A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (30-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 141899)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Furo,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., Larocque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meidrim,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nanyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigillo,J., Vassiliev,H.,
Viel,R., Vo.A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (05-JAN-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 141899)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Furo,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., Larocque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meidrim,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nanyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigillo,J., Vassiliev,H.,
Viel,R., Vo.A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission

```

COMMENT

Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Nov 30, 2001 this sequence version replaced gi:15528911.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RW/RepeatMasker.html  
 Only the first 141.9 kilobases of this clone are being submitted.  
 The remainder overlaps accession number AC018868.  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L11876  
 Center clone name: 814\_P\_5  
 -----

FEATURES

source  
 1. .141899  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="15"  
 /map="15"  
 /clone="RP11-814P5"  
 /clone\_lib="RPC1-11 Human Male BAC"  
 1. .274  
 /notes="single clone coverage"  
 1. .15  
 /note="<30 qual single clone coverage"  
 30. .35  
 /note="<30 qual single clone coverage"  
 293. .298  
 /note="<30 qual SINGL region"  
 382. .396  
 /note="<30 qual SINGL region"  
 412. .416  
 /note="<30 qual SINGL region"  
 428. .432  
 /note="<30 qual SINGL region"  
 487. .575  
 /note="single clone coverage"  
 494. .498  
 /note="<30 qual single clone coverage"  
 528. .532  
 /note="<30 qual single clone coverage"  
 541. .545  
 /note="<30 qual single clone coverage"  
 586. .590  
 /note="<30 qual SINGL region"  
 1021. .1530  
 /rpt\_family="MER34B"  
 complement(1726..2118)  
 /rpt\_family="MER57A"  
 complement(2570..2604)  
 /rpt\_family="L2"  
 2770. .3093  
 /rpt\_family="ALuY"  
 3317. .3612  
 /rpt\_family="ALuJo"  
 complement(4209..4961)  
 /rpt\_family="L1MA4"  
 6239. .6415  
 /rpt\_family="MIR"  
 7440. .7484  
 /rpt\_family="(TCTCC)n"  
 7570. .7782  
 /rpt\_family="MIR"  
 complement(7800..7896)  
 /rpt\_family="MIR"  
 7944. .8436  
 /rpt\_family="MLT1K"  
 complement(8841..9063)  
 /rpt\_family="MIR"

unsure  
 complement(8992..9086)  
 /note="single clone coverage"  
 unsure  
 complement(8996..9001)  
 /note="<30 qual single clone coverage"  
 unsure  
 complement(9035..9039)  
 /note="<30 qual single clone coverage"  
 unsure  
 complement(9052..9063)  
 /note="<30 qual single clone coverage"  
 unsure  
 complement(9120..9137)  
 /note="<30 qual SINGL region"  
 repeat\_region  
 complement(9179..9316)  
 /rpt\_family="FAM"  
 9436. .9567  
 /rpt\_family="MIR"  
 9570. .9603  
 /rpt\_family="(CAT)n"  
 9729. .9756  
 /rpt\_family="(CAAAA)n"  
 complement(10861..10957)  
 /rpt\_family="HAL1b"  
 repeat\_region  
 complement(11110..11325)  
 /rpt\_family="LIME"  
 repeat\_region  
 complement(11327..11603)  
 /rpt\_family="AluJo"  
 11447. .11452  
 /note="<30 qual SINGL region"  
 11457. .11461  
 /note="<30 qual SINGL region"  
 repeat\_region  
 complement(11728..11816)  
 /rpt\_family="MIR3"  
 repeat\_region  
 complement(11822..12123)  
 /rpt\_family="AluSx"  
 12056. .12061  
 /note="<30 qual SINGL region"  
 12129. .12136  
 /note="<30 qual SINGL region"  
 12168. .12174  
 /note="<30 qual SINGL region"

Alignment Scores:

Pred. No.: 663 Length: 141899  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 4.55% Indels: 0  
 DB: 9 Gaps: 0

US-09-966-880A-8 (1-198) x AC087457 (1-141899)

QY 97 PheLeuArgGlyAsnProAsnLeuSer 105

Db 62431 TTCTTAAGGGGCAACCCCAACCTGTCT 62457

RESULT 60

AC123206

LOCUS

DEFINITION

\*\*\* 74 unordered pieces.

AC123206

AC123206.2 GI:21671690

HTG; HTGS\_PHASE1.

Norway rat.

SOURCE

ORGANISM

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

REFERENCE

1 (bases 1 to 150147)

AUTHORS

Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,

Alsbrooks, S.L., Amaratunge, H.C., Are, J.R., Ayele, M., Banks, T.,

Barbata, J., Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D.,

Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,

Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,

Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,

AC123206 150147 bp DNA linear HTG 17-JUL-2002  
 Rattus norvegicus clone CH230-137M23, \*\*\* SEQUENCING IN PROGRESS

Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,  
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,  
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,  
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,  
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,  
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,  
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,  
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,  
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,  
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,  
Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,  
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,  
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,  
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,  
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louiseged,H.,  
Lozador,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,  
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,  
Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,  
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,  
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,  
Nguyen,N., Nickerson,E., Nwckenkwo,S., Oguh,M., Okwuonu,G.,  
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,D., Perez,L.,  
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,  
Rives,M., Rojas,A., Rojibokan,I., Roife,M., Ruiz,S., Savery,G.,  
Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,  
Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H.,  
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,  
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,  
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,  
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,  
Williams,G., Williamson,A., Wleczkyk,R., Wooden,S., Worley,K.,  
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
Weinstock,G. and Gibbs,R.

Direct Submission  
Unpublished  
2 (bases 1 to 150147)  
Worley,K.C.

Direct Submission  
Submitted (29-MAY-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 150147)  
Worley,K.C.

Direct Submission  
Submitted (17-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

On Jul 2, 2002 this sequence version replaced gi:21239988.

----- Genome Center  
Center: Baylor College of Medicine  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu

----- Project Information  
Center project name: GFGZ  
Center clone name: CH230-137M23  
----- Summary Statistics  
Sequencing vector: Plasmid;  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 85843 bases at least Q40  
Consensus quality: 93252 bases at least Q30  
Consensus quality: 97242 bases at least Q20  
-----

\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 74 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number: will  
\* be preserved.

1 1454: contig of 1454 bp in length  
1455: gap of unknown length  
1555: contig of 1679 bp in length  
3234: gap of unknown length  
3334: contig of 1128 bp in length  
4461: gap of unknown length  
4561: gap of unknown length  
5979: contig of 1418 bp in length  
6079: gap of unknown length  
7681: contig of 1602 bp in length  
7682: gap of unknown length  
7781: gap of unknown length  
9391: contig of 1610 bp in length  
9392: gap of unknown length  
9491: contig of 1233 bp in length  
1024: gap of unknown length  
10825: gap of unknown length  
10825: contig of 1177 bp in length  
12001: gap of unknown length  
12002: gap of unknown length  
12101: gap of unknown length  
13102: contig of 1460 bp in length  
13561: gap of unknown length  
13662: gap of unknown length  
14708: contig of 1047 bp in length  
14808: gap of unknown length  
14809: gap of unknown length  
15880: contig of 1072 bp in length  
15881: gap of unknown length  
15980: gap of unknown length  
17340: contig of 1360 bp in length  
17341: gap of unknown length  
17440: gap of unknown length  
19078: contig of 1638 bp in length  
19178: gap of unknown length  
19179: gap of unknown length  
20869: contig of 1691 bp in length  
20870: gap of unknown length  
20969: gap of unknown length  
22724: contig of 1755 bp in length  
22725: gap of unknown length  
22824: gap of unknown length  
23923: contig of 1099 bp in length  
24023: gap of unknown length  
24024: gap of unknown length  
25106: contig of 1083 bp in length  
25206: gap of unknown length  
25207: contig of 1353 bp in length  
26559: contig of 1353 bp in length  
26560: gap of unknown length  
27226: contig of 1067 bp in length  
27227: gap of unknown length  
27826: gap of unknown length  
27827: gap of unknown length  
28885: contig of 1059 bp in length  
28886: gap of unknown length  
28986: gap of unknown length  
30877: contig of 1892 bp in length  
30878: gap of unknown length  
30978: gap of unknown length  
32206: contig of 1229 bp in length  
32207: gap of unknown length  
33207: gap of unknown length  
33895: contig of 1589 bp in length  
33995: gap of unknown length  
35594: contig of 1599 bp in length  
35595: gap of unknown length  
35694: gap of unknown length  
37941: contig of 2247 bp in length  
37942: gap of unknown length  
38041: gap of unknown length  
38042: contig of 1903 bp in length  
39945: gap of unknown length  
40044: gap of unknown length  
41049: contig of 1005 bp in length  
41149: gap of unknown length  
41150: gap of unknown length  
41150: contig of 1426 bp in length  
42575: gap of unknown length  
42576: gap of unknown length  
42576: gap of unknown length  
43884: contig of 1209 bp in length  
43885: gap of unknown length  
43885: gap of unknown length  
45774: contig of 1790 bp in length  
45775: gap of unknown length  
45874: gap of unknown length  
45875: contig of 1554 bp in length  
47429: gap of unknown length  
47528: gap of unknown length  
4819: contig of 1291 bp in length  
4819: gap of unknown length  
48920: contig of 1398 bp in length  
50117: gap of unknown length  
50117: gap of unknown length  
50418: contig of 1145 bp in length  
51562: gap of unknown length  
51563: gap of unknown length  
52781: contig of 1119 bp in length  
52881: gap of unknown length  
52882: contig of 1768 bp in length



```

* 54650 54749: gap of unknown length
* 54750 5624: contig of 1775 bp in length
* 5625 56624: gap of unknown length
* 56625 57872: contig of 1248 bp in length
* 57873 57972: gap of unknown length
* 57973 59161: contig of 1195 bp in length
* 59168 59267: gap of unknown length
* 59268 60716: contig of 1449 bp in length
* 60717 60816: gap of unknown length
* 60817 63158: contig of 2342 bp in length
* 63159 63258: gap of unknown length
* 63259 63761: contig of 2503 bp in length
* 63762 65861: gap of unknown length
* 65862 67874: contig of 2013 bp in length
* 67875 67974: gap of unknown length
* 67975 69539: contig of 1565 bp in length
* 69540 69639: gap of unknown length
* 69640 72055: contig of 2416 bp in length
* 72056 72155: gap of unknown length
* 72156 74404: contig of 2249 bp in length
* 74405 74504: gap of unknown length
* 74505 75761: contig of 1257 bp in length
* 75762 75861: gap of unknown length
* 75862 78235: contig of 2374 bp in length
* 78236 78335: gap of unknown length
* 78336 79893: contig of 1558 bp in length
* 79894 82644: contig of 2651 bp in length
* 82645 82744: gap of unknown length
* 82745 84805: contig of 2061 bp in length
* 84806 84905: gap of unknown length
* 84906 86751: contig of 1846 bp in length
* 86752 86851: gap of unknown length
* 86852 88927: contig of 2076 bp in length
* 88928 89027: gap of unknown length
* 89028 91403: contig of 2376 bp in length

```

## Alignment Scores:

```

Pred. No.: 696 Length: 150147
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.53% Indels: 0
DB: 2 Gaps: 0

```

US-09-966-880A-8 (1-198) x AC123206 (1-150147)

```

Qy 42 PheSerLeuAspPheGlyTyrLeuArg 50
|||||TTTTTTTTTTTTTTTTTTTTTTTT
Db 69791 TTTCTTGGATTTTGGTTATTACGG 69817

```

```

RESULT 61
AL513164
LOCUS AL513164 154288 bp DNA linear PRI 26-JUN-2001
DEFINITION Human DNA sequence from clone RP11-400P17 on chromosome 6, complete
sequence.
ACCESSION AL513164
VERSION AL513164.13 GI:14572569
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 154288)
Direct Submission
AUTHORS Sanger Centre, Hinxton, Cambridgeshire,
TITLE Submitted (26-JUN-2001) humquery@sanger.ac.uk
JOURNAL CB10 ISA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jun 27, 2001 this sequence version replaced gi:14268219.
COMMENT During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the

```

variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; information on the WORMPEP database can be found at

<http://www.sanger.ac.uk/Projects/C/elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr6> RP11-400P17 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pBACe3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-400P17 It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.

The true right end of clone RP11-400P17 is at 154288 in this sequence. The true left end of clone RP11-69H11 is at 28928 in this sequence. The true right end of clone RP11-631F7 is at 2000 in this sequence.

## FEATURES

## Location/Qualifiers

```

source
1..154288
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RP11-400P17"
/clone_lib="RPCI-11.2"
543..849
/note="AluY repeat: matches 1..311 of consensus"
1705..1828
/note="L2 repeat: matches 2610..2750 of consensus"
2455..2596
/note="MER69 repeat: matches 43..2510 of consensus"
3862..3918
/note="L2 repeat: matches 2694..2750 of consensus"
4140..4183
/note="22 copies 2 mer tg 79% conserved"
4146..4181
/note="9 copies 4 mer tgtg 83% conserved"
4344..4551
/note="MIR repeat: matches 20..241 of consensus"
4522..4664
/note="MIR repeat: matches 80..235 of consensus"
4720..4776
/note="MIR repeat: matches 165..221 of consensus"
4767..4826
/note="MIR repeat: matches 195..257 of consensus"
4776..4837
/note="L2 repeat: matches 2689..2750 of consensus"
4840..4943
/note="MIR repeat: matches 17..147 of consensus"
5410..5465
/note="28 copies 2 mer aa 71% conserved"
6529..6810
/note="L1P3 repeat: matches 5864..6146 of consensus"
7313..7615
/note="AluSx repeat: matches 1..303 of consensus"
7946..8427
/note="L1MB6 repeat: matches 4666..5136 of consensus"
8428..8724
/note="AluSc repeat: matches 1..296 of consensus"
8725..8760
/note="L1MB6 repeat: matches 5136..5170 of consensus"

```

```
repeat_region 8761..9067
/note="Alusq repeat: matches 1..306 of consensus"
repeat_region 9068..9174
/note="L1MB6 repeat: matches 5170..5305 of consensus"
repeat_region 9193..9350
/note="FAM repeat: matches 134..43 of consensus"
repeat_region 9353..9969
/note="L1MB6 repeat: matches 5311..5919 of consensus"
repeat_region 9971..10354
/note="MLR1B repeat: matches 1..388 of consensus"
repeat_region 10467..10778
/note="AluJb repeat: matches 1..308 of consensus"
repeat_region 11491..11674
/note="L1MB1 repeat: matches 5955..6163 of consensus"
repeat_region 11708..11915
/note="L1MB8 repeat: matches 5938..6166 of consensus"
repeat_region 12438..12465
/note="L2 repeat: matches 2579..2604 of consensus"
repeat_region 12466..12764
/note="AlusC repeat: matches 1..300 of consensus"
repeat_region 12765..12905
/note="L2 repeat: matches 2604..2750 of consensus"
repeat_region 12921..13018
/note="FLAM_A repeat: matches 37..133 of consensus"
repeat_region 13172..13375
/note="MIR repeat: matches 53..262 of consensus"
repeat_region 14711..14894
/note="FRAM repeat: matches 6..174 of consensus"
repeat_region 15163..15460
/note="AluJo repeat: matches 1..300 of consensus"
repeat_region 15634..15940
/note="AlusX repeat: matches 2..305 of consensus"
repeat_region 16932..16959
/note="L4 copies 2 mer ca 92% conserved"
repeat_region 17439..17737
/note="Alusq repeat: matches 1..310 of consensus"
repeat_region 19260..19611
/note="L2 repeat: matches 2102..2471 of consensus"
repeat_region 20009..20171
/note="FRAM repeat: matches 9..173 of consensus"
repeat_region 20172..20223
/note="L3 copies 4 mer gggg 90% conserved"
repeat_region 20810..21105
/note="Alusq repeat: matches 1..300 of consensus"
repeat_region 21217..21351
/note="MIR repeat: matches 84..221 of consensus"
repeat_region 21442..21747
/note="AluJb repeat: matches 1..297 of consensus"
repeat_region 21893..22005
/note="MIR repeat: matches 96..212 of consensus"
repeat_region 22080..22198
/note="MER5B repeat: matches 1..126 of consensus"
repeat_region 22201..22232
/note="L6 copies 2 mer tg 96% conserved"
repeat_region 22203..22230
/note="L7 copies 4 mer ttgtg 100% conserved"
repeat_region 22233..22514
/note="AluY repeat: matches 1..282 of consensus"
repeat_region 22525..22575
/note="MER5B repeat: matches 125..175 of consensus"
repeat_region 24124..24384
/note="AluJb repeat: matches 5..302 of consensus"
repeat_region 24472..24620
/note="MLR1B repeat: matches 1..151 of consensus"
repeat_region 24987..25118
/note="MIR repeat: matches 1..137 of consensus"
repeat_region 25742..26046
/note="Alusq repeat: matches 1..305 of consensus"
repeat_region 27815..28096
/note="AluJo repeat: matches 1..282 of consensus"
repeat_region 28120..28430
/note="AlusX repeat: matches 1..311 of consensus"
repeat_region 28638..28718
```

```
/note="L2 repeat: matches 2663..2745 of consensus"
30584..30857
/note="AluJo repeat: matches 1..275 of consensus"
30892..31125
/note="MIR repeat: matches 12..255 of consensus"
31641..31763
/note="FLAM_A repeat: matches 1..122 of consensus"
31791..32009
/note="MIR repeat: matches 18..255 of consensus"
32551..32851
/note="Alusg repeat: matches 1..301 of consensus"
32866..33173
/note="AlusX repeat: matches 1..312 of consensus"
33511..33615
/note="U6 repeat: matches 3..107 of consensus"
33831..34127
/note="AlusX repeat: matches 1..295 of consensus"
34169..34263
/note="MIR repeat: matches 159..247 of consensus"
34222..34281
/note="L2 repeat: matches 2692..2747 of consensus"
34351..34836
/note="AlusX repeat: matches 2..297 of consensus"
34959..35034
/note="L2 repeat: matches 2630..2707 of consensus"
35209..35340
/note="MIR repeat: matches 112..254 of consensus"
35421..35662
/note="AluJo repeat: matches 57..300 of consensus"
35663..35692
/note="L5 copies 2 mer ga 86% conserved"
35772..36094
/note="AlusX repeat: matches 1..309 of consensus"
36095..36335
```

Alignment Scores:  
Pred. No.: 713 Length: 154288  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 4.55% Indels: 0  
DB: 9 Gaps: 0

US-09-966-880A-8 (1-198) x ALS13164 (1-154288)  
QY 125 GlyLeuArgArgLeuHisArgAlaGly 133  
Db 18890 GGCTTAAGGAGTTACATAGGCTGGA 18916

RESULT 62  
AC113662 158016 bp DNA linear HTG 13-JUL-2002  
LOCUS Rattus norvegicus clone CH230-15401, \*\*\* SEQUENCING IN PROGRESS  
DEFINITION \*\*\* 56 unordered pieces.  
AC113662  
VERSION AC113662.3 GI:21743470  
KEYWORDS HTG: HTGS\_PHASE1.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE 1 (bases 1 to 158016)  
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,  
Alsbrooks,S.L., AmaralTunge,H.C., Are,J.R., Ayale,M., Banks,T.,  
Barbora,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,  
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,  
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,  
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,  
Cleveland,C.D., Cox,C., Coyle,M.B., Dathorne,S.R., David,R.,  
Devilla,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,  
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,

```

Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Haviak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,
Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlisson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarger,O., Lieu,C., Liu,J., Liu,W., Louised,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mahoney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,S., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
Oraguwe,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shoostari,N., Sisson,I.,
Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 158016)
Worley,K.C.
Direct Submission
Submitted (05-MAR-2002) Human Genome Sequencing Center, Department
Of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 158016)
Worley,K.C.
Direct Submission
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
Of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 13, 2002 this sequence version replaced gi:19525824.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GRUR
Center clone name: CH230-15401
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 116208 bases at least Q40
Consensus quality: 119847 bases at least Q30
Consensus quality: 122552 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 56 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1728: contig of 1728 bp in length
* 1729 1828: gap of unknown length

```

```

1829 3040: contig of 1212 bp in length
3041 3140: gap of unknown length
3141 4238: contig of 1098 bp in length
4239 4338: gap of unknown length
4339 5474: contig of 1136 bp in length
5475 5574: gap of unknown length
5575 6934: contig of 1360 bp in length
6935 7034: gap of unknown length
7035 8042: contig of 1008 bp in length
8043 8142: gap of unknown length
8143 9327: contig of 1185 bp in length
9328 9427: gap of unknown length
9428 10467: contig of 1040 bp in length
10468 10567: gap of unknown length
10568 11660: contig of 1093 bp in length
11661 11760: gap of unknown length
11761 13009: contig of 1249 bp in length
13010 13109: gap of unknown length
13110 14454: contig of 1345 bp in length
14455 14554: gap of unknown length
14555 15646: contig of 1092 bp in length
15647 15746: gap of unknown length
15747 16820: contig of 1074 bp in length
16821 16920: gap of unknown length
16921 18633: contig of 1713 bp in length
18634 18733: gap of unknown length
18734 20290: contig of 1557 bp in length
20291 20390: gap of unknown length
20391 21752: contig of 1362 bp in length
21753 21852: gap of unknown length
21853 22933: contig of 1081 bp in length
22934 23033: gap of unknown length
23034 25044: contig of 2011 bp in length
25045 25144: gap of unknown length
25145 26595: contig of 1451 bp in length
26596 26695: gap of unknown length
26696 28801: contig of 2106 bp in length
28802 28901: gap of unknown length
28902 30135: contig of 1234 bp in length
30136 30235: gap of unknown length
30236 32675: contig of 2440 bp in length
32676 34111: contig of 1336 bp in length
34112 34211: gap of unknown length
34212 36075: contig of 1864 bp in length
36076 36175: gap of unknown length
36176 38430: contig of 2255 bp in length
38431 38530: gap of unknown length
38531 40640: contig of 2110 bp in length
40641 40740: gap of unknown length
40741 43028: contig of 2288 bp in length
43029 43128: gap of unknown length
43129 45507: contig of 2379 bp in length
45508 45607: gap of unknown length
45608 47252: contig of 1645 bp in length
47253 47352: gap of unknown length
47353 48961: contig of 1609 bp in length
48962 49061: gap of unknown length
49062 51776: contig of 2715 bp in length
51777 51876: gap of unknown length
51877 53989: contig of 2113 bp in length
53990 54099: gap of unknown length
54099 56499: contig of 2410 bp in length
56500 56599: gap of unknown length
56600 59667: contig of 3068 bp in length
59668 59767: gap of unknown length
59768 62514: contig of 2747 bp in length
62515 62614: gap of unknown length
62615 64470: contig of 1856 bp in length
64471 64570: gap of unknown length
64571 66820: contig of 2250 bp in length
66821 66920: gap of unknown length
66921 68862: contig of 1942 bp in length

```

\* 68963 68962: gap of unknown length  
 \* 68963 contig of 2866 bp in length  
 \* 71928 71928: gap of unknown length  
 \* 71929 71928: contig of 2820 bp in length  
 \* 74749 74748: gap of unknown length  
 \* 74849 74849: contig of 3211 bp in length  
 \* 78059 78059: gap of unknown length  
 \* 78159 78159: contig of 4097 bp in length  
 \* 82256 82256: gap of unknown length  
 \* 82257 82257: contig of 3219 bp in length  
 \* 82357 82357: gap of unknown length  
 \* 85675 85675: contig of 3657 bp in length  
 \* 85676 85676: gap of unknown length  
 \* 89332 89332: gap of unknown length  
 \* 89333 89332: contig of 4421 bp in length  
 \* 93853 93853: gap of unknown length  
 \* 93854 93853: contig of 3779 bp in length  
 \* 93954 93954: gap of unknown length  
 \* 97732 97732: contig of 4531 bp in length  
 \* 97733 97732: gap of unknown length  
 \* 102363 102363: contig of 4531 bp in length  
 \* 102364 102363: gap of unknown length  
 \* 102464 102464: contig of 6553 bp in length  
 \* 109017 109017: gap of unknown length  
 \* 109117 109117: contig of 6082 bp in length  
 \* 115199 115199: gap of unknown length  
 \* 115299 115299: contig of 5495 bp in length  
 \* 120794 120793: gap of unknown length  
 \* 120894 120893: contig of 4445 bp in length  
 \* 125339 125338: gap of unknown length  
 \* 125439 125438: contig of 6067 bp in length  
 \* 131506 131505: gap of unknown length  
 \* 131606 131605: gap of unknown length  
 \* 138366 138366: contig of 6761 bp in length  
 \* 138367 138366: gap of unknown length  
 \* 138467 144803: contig of 6337 bp in length

Alignment Scores:  
 Pred. No.: 728 Length: 158016  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 4.55% Indels: 0  
 DB: 2 Gaps: 0

US-09-966-880A-8 (1-198) x AC113662 (1-158016)

Qy 175 GlnLeuArgArgIleLeuLeuProLeu 183  
 Db 151669 CAGCTACGTAGAAATTTACTCTTGT 151695

RESULT 63  
 AP002436/c  
 LOCUS Homo sapiens genomic DNA, chromosome 11q clone:RP11-627G1, complete sequences.  
 DEFINITION AP002436.3 GI:13094217  
 ACCESSION AP002436  
 VERSION AP002436.3  
 KEYWORDS HTG.  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (sites)  
 AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.  
 TITLE Published Only in Database (2000)  
 JOURNAL 2 (bases 1 to 173509)  
 REFERENCE Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.  
 AUTHORS Direct Submission  
 TITLE Submitted (02-JUN-2000) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 JOURNAL (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)

COMMENT On Feb 21, 2001 this sequence version replaced gi:11071943.  
 FEATURES Location/Qualifiers  
 source 1. 173509  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="11"  
 /map="11q"  
 /clone="RP11-627G1"  
 BASE COUNT 53330 a 37471 c 36148 g 46560 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 789 Length: 173509  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 4.55% Indels: 0  
 DB: 9 Gaps: 0  
 US-09-966-880A-8 (1-198) x AP002436 (1-173509)  
 Qy 39 AlaThrSerPheSerLeuAspPheGly 47  
 Db 82024 GCACCACTTCTCTCTAGATTTGGT 81998  
 RESULT 64  
 AC068190  
 LOCUS Homo sapiens chromosome 11 clone RP11-419E16 map 11, WORKING DRAFT  
 DEFINITION AC068190  
 ACCESSION AC068190.2 GI:8347999  
 VERSION AC068190  
 KEYWORDS HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 176053)  
 AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
 TITLE Homo sapiens chromosome 11, clone RP11-419E16  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 176053)  
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bedalov,F., Boguslavsky,L., Bouckhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hago,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatus,A., Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoucq,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheters,R., Meldrum,J., Meneus,L., Mihoval,T., Miranda,C., Mlenga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trifilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission  
 Submitted (29-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 JOURNAL 3 (bases 1 to 176053)  
 REFERENCE 3 (bases 1 to 176053)  
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Bouckhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faros,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,

Grand-pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,  
 Iliev,I., Johnson,R., Jones,C., Karatas,A., Laroque,K.,  
 Lanazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,  
 MacLean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M.,  
 McEwan,P., McKenna,K., McPheeters,R., Meldrim,J., Meneus,L.,  
 Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,  
 Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,  
 Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,  
 Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,  
 Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,  
 Severy,P., Sougnuez,C., Spencer,B., Stange-Thomann,N.,  
 Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,  
 Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,  
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,N.

# TITLE JOURNAL

## COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WtBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence.submissions@genome.wi.mit.edu](mailto:sequence.submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L9913

Center clone name: 419\_E\_16

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 170544 bases at least Q40

Consensus quality: 173410 bases at least Q30

Consensus quality: 174372 bases at least Q20

Insert size: 176000; agarose-fp

Quality coverage: 5.0 in Q20 bases; agarose-fp

Quality coverage: 5.1 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 10 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1 1361: contig of 1361 bp in length  
 \* 1362 1461: gap of 100 bp  
 \* 1462 4481: contig of 3020 bp in length  
 \* 4482 4581: gap of 100 bp  
 \* 4582 12932: contig of 8351 bp in length  
 \* 12933 13032: gap of 100 bp  
 \* 13033 27303: contig of 14271 bp in length  
 \* 27304 27403: gap of 100 bp  
 \* 27404 39940: contig of 12537 bp in length  
 \* 39941 40040: gap of 100 bp  
 \* 40041 60017: contig of 19977 bp in length  
 \* 60018 60117: gap of 100 bp  
 \* 60118 83591: contig of 23474 bp in length  
 \* 83592 83691: gap of 100 bp  
 \* 83692 108521: contig of 24830 bp in length  
 \* 108522 108621: gap of 100 bp  
 \* 108622 139440: contig of 30819 bp in length  
 \* 139441 139540: gap of 100 bp  
 \* 139541 176053: contig of 36513 bp in length.  
 \* Location/Qualifiers  
 \* 1. 176053.  
 \* /organism="Homo sapiens"  
 \* /db\_xref="taxon:9606"

## FEATURES source

/chromosome="11"  
 /map="11"  
 /clones="RP11-419E16"  
 /clone\_lib="RP11 Human Male BAC"  
 1. 1361  
 /note="assembly\_fragment"  
 1462. 4481  
 /note="assembly\_fragment"  
 4582. 12932  
 /note="assembly\_fragment"  
 13033. 27303  
 /note="assembly\_fragment"  
 clone\_end:T7  
 vector\_side:left  
 27404. 39940  
 /note="assembly\_fragment"  
 40041. 60017  
 /note="assembly\_fragment"  
 60118. 83591  
 /note="assembly\_fragment"  
 83692. 108521  
 /note="assembly\_fragment"  
 108622. 139440  
 /note="assembly\_fragment"  
 139541. 176053  
 /note="assembly\_fragment"  
 clone\_end:SP6  
 vector\_side:left  
 49213 a 37507 c 38347 g 50081 t 905 others  
 BASE COUNT  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 799 Length: 176053  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 4.53% Indels: 0  
 DB: 2 Gaps: 0  
 US-09-966-880A-8 (1-198) x AC068190 (1-176053)  
 Qy 39 AlathrSerPheSerLeuAspPheGly 47  
 Db 117796 GCAACCAAGTTCTCTAGATTTTGGT 117822  
 RESULT 65  
 AC127421 179510 bp DNA linear HTG 17-JUL-2002  
 LOCUS AC127421 Mus musculus chromosome UNK clone RP24-314F1, WORKING DRAFT  
 DEFINITION AC127421  
 ACCESSION AC127421  
 VERSION AC127421.1 GI:21759528  
 KEYWORDS HTG; HTGS\_PHRASE1; HTGS\_DRAFT.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 1 (bases 1 to 179510)  
 McPherson,J.D. and Waterston,R.H.  
 McPherson,J.D. and Waterston,R.H.  
 The sequence of Mus musculus clone  
 2 (bases 1 to 179510)  
 McPherson,J.D. and Waterston,R.H.  
 Direct Submission  
 Submitted (15-JUL-2002) Genome Sequencing Center, 4444 Forest Park  
 Parkway, St. Louis, MO 63108, USA  
 3 (bases 1 to 179510)  
 McPherson,J.D. and Waterston,R.H.  
 Direct Submission  
 Submitted (17-JUL-2002) Genome Sequencing Center, 4444 Forest Park  
 Parkway, St. Louis, MO 63108, USA  
 ----- Genome Center -----  
 COMMENT

Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc/index.shtml>  
Contact: [submissions@watson.wustl.edu](mailto:submissions@watson.wustl.edu)  
----- Project Information -----  
Center project name: M\_BB0314F01  
-----

----- Summary Statistics -----  
Sequencing vector: M13; 0%  
Chemistry: Dye-terminator; plasmid; 100%  
Assembly: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 181162 bases at least Q40  
Consensus quality: 181808 bases at least Q30  
Consensus quality: 182185 bases at least Q20  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 8 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1378: contig of 1378 bp in length  
1379 1478: gap of unknown length  
1479 4229: contig of 2751 bp in length  
4230 4329: gap of unknown length  
4330 10616: contig of 6287 bp in length  
10617 10716: gap of unknown length  
10717 34165: contig of 23449 bp in length  
34166 34265: gap of unknown length  
34266 55394: gap of unknown length  
55395 54949: gap of unknown length  
54950 92891: contig of 37397 bp in length  
92892 129685: contig of 36694 bp in length  
129686 129785: gap of unknown length  
129786 179510: contig of 49725 bp in length.

FEATURES  
source

1..179510  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="UNK"  
/clones="RP24-314F1"

misc\_feature 1..1378  
/note="assembly\_name:Contig17"  
misc\_feature 1479..4229  
/note="assembly\_name:Contig18"  
misc\_feature 4330..10616  
/note="assembly\_name:Contig20"  
misc\_feature 10717..34165  
/note="assembly\_name:Contig21"  
misc\_feature 34266..55394  
/note="assembly\_name:Contig22"  
misc\_feature 55495..92891  
/note="assembly\_name:Contig23"  
misc\_feature 92992..129685  
/note="assembly\_name:Contig24"  
misc\_feature 129786..179510  
/note="assembly\_name:Contig25"  
BASE COUNT 51663 a 37702 c 36620 g 52806 t 719 others  
ORIGIN

Alignment Scores:

Pred. No.: 812 Length: 179510  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 4.55% Indels: 0  
DB: 2 Gaps: 0

US-09-966-880A-8 (1-198) x AC127421 (1-179510)  
Oy 103 AsnLeuSerLeuArgIlePheThrAla 111  
|||||  
Db 63336 AACCTCTCTCTTCGCATATTACTGCT 63362

RESULT 66

AC020857  
LOCUS AC020857 180668 bp DNA linear HTG 16-FEB-2000  
DEFINITION Mus musculus clone RP23-163F21, LOW-PASS SEQUENCE SAMPLING.  
AC020857  
VERSION AC020857.2 GI:6980198  
KEYWORDS HTG; HTGS PHASE0.  
SOURCE Mus musculus.  
ORGANISM Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

COMMENT

On Feb 16, 2000 this sequence version replaced gi:6686451.  
\* NOTE: This record contains 205 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 796: contig of 796 bp in length  
gap of unknown length  
797 1094: contig of 298 bp in length  
gap of unknown length  
1095 2104: contig of 1010 bp in length  
gap of unknown length  
2105 3041: contig of 937 bp in length  
gap of unknown length  
3042 3673: contig of 632 bp in length  
gap of unknown length  
3674 4251: contig of 578 bp in length  
gap of unknown length  
4252 4677: contig of 426 bp in length  
gap of unknown length  
4678 5588: contig of 911 bp in length  
gap of unknown length  
5589 6457: contig of 869 bp in length  
gap of unknown length  
6458 6775: contig of 318 bp in length  
gap of unknown length  
6776 7759: contig of 984 bp in length  
gap of unknown length  
7760 8466: contig of 707 bp in length  
gap of unknown length  
8467 8733: contig of 267 bp in length  
gap of unknown length  
8734 9823: contig of 1090 bp in length  
gap of unknown length  
9824 10617: contig of 794 bp in length  
gap of unknown length  
10618 11286: contig of 669 bp in length  
gap of unknown length  
11287 11422: contig of 136 bp in length  
gap of unknown length  
11423 11683: contig of 261 bp in length  
gap of unknown length



Qy 103 AsnLeuSerLeuArgIlePheThrAla 111  
 |||||  
 Db 146416 AACCTCTCTCTCGGATATTTACTGCT 146442

RESULT 67  
 AP003402c 187948 bp DNA linear PRI 06-NOV-2001  
 DEFINITION Homo sapiens genomic DNA, chromosome 11q clone:RP11-661121,  
 complete sequence.  
 ACCESSION AP003402  
 VERSION AP003402.2 GI:16751506  
 KEYWORDS HTG.  
 SOURCE Homo sapiens DNA, clone:RP11-661121.  
 ORGANISM Homo sapiens

REFERENCE  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1

AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,  
 Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.

TITLE Homo sapiens genomic DNA  
 JOURNAL Published Only in Database (2001)  
 REFERENCE 2 (bases 1 to 187948)  
 AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,  
 Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.

TITLE Direct Submission  
 JOURNAL Submitted (13-MAR-2001) Masahira Hattori, The Institute of Physical  
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC),  
 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 (E-mail:hattori@psc.riken.go.jp, URL:http://hgsc.riken.go.jp/  
 Tel:81-45-503-9111, Fax:81-45-503-9170)  
 On Nov 5, 2001 this sequence version replaced gi:13366105.

COMMENT Location/Qualifiers  
 FEATURES  
 source  
 1..187948  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /map="11q"  
 /clone="RP11-661121"  
 BASE COUNT 55717 a 41857 c 40668 g 49706 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 845 Length: 187948  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 4.55% Indels: 0  
 DB: 9 Gaps: 0

US-09-966-880A-8 (1-198) x AP003402 (1-187948)  
 Qy 39 AlatrSerPheSerLeuAspPheGly 47  
 |||||  
 Db 159621 GCACACAGTTCTCTCTAGATTTTGGT 159595

RESULT 68  
 AC007083 188067 bp DNA linear INV 17-MAR-2001  
 LOCUS Drosophila melanogaster, chromosome 2L, region 33A-, BAC clone  
 DEFINITION BACR43K24, complete sequence.  
 ACCESSION AC007083  
 VERSION AC007083.11 GI:13374638  
 KEYWORDS HTG.  
 SOURCE Drosophila melanogaster.  
 ORGANISM Drosophila melanogaster

REFERENCE  
 Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Ephydroidea; Drosophilidae; Drosophila.  
 1 (bases 1 to 188067)  
 Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,  
 Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,  
 Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,  
 Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,

Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,  
 Ferreira,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,  
 Gonzalez,M., Houch,J., Hoskins,R.A., Hostin,D., Howland,T.J.,  
 Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,  
 McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,  
 Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,  
 Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,  
 Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,  
 Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.  
 Sequencing of Drosophila chromosome 2L, region 33A-  
 Unpublished  
 2 (bases 1 to 188067)  
 Celniker,S.E., Aghayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,  
 Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,  
 Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,  
 Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,  
 Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Madxa,P.,  
 Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,  
 Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,  
 Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zierun,L.L. and  
 Rubin,G.M.  
 Direct Submission  
 Submitted (16-MAR-1999) Drosophila Genome Center, Lawrence Berkeley  
 Laboratory, MS 64-121, Berkeley, CA 94720, USA  
 On Mar 17, 2001 this sequence version replaced gi:7264773.  
 Sequence submitted by:  
 Berkeley Drosophila Genome Project  
 Lawrence Berkeley National Laboratory, MS 64-121  
 Berkeley, CA 94720  
 This sequence was assembled using end sequences from a whole genome  
 shotgun and from subclones of this BAC and its neighboring clones.  
 For further information about this sequence, including its location  
 and relationship to other sequences, please visit our sequence  
 archive Web site (<http://www.fruitfly.org/sequence/>) or send email  
 to [bdgpf@fruitfly.berkeley.edu](mailto:bdgpf@fruitfly.berkeley.edu).

FEATURES  
 Location/Qualifiers  
 1..188067  
 /organism="Drosophila melanogaster"  
 /strain="y: cn bw sp"  
 /db\_xref="taxon:7227"  
 /chromosome="2L"  
 /map="33A-"  
 /clone="BACR43K24 (D575)"  
 /clone\_lib="RPC1-98 (Roswell Park Cancer Institute  
 Drosophila melanogaster BAC library, partial ECORI in  
 pBACE3.6)"  
 BASE COUNT 53731 a 40263 c 40250 g 53823 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 846 Length: 188067  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 4.55% Indels: 0  
 DB: 3 Gaps: 0

US-09-966-880A-8 (1-198) x AC007083 (1-188067)  
 Qy 117 GluAspArgLysAlaGluProGluGly 125  
 |||||  
 Db 105148 GAGGACAGAAAGGAGAGAGAGGGT 105174

RESULT 69  
 AL840638 188136 bp DNA linear HTG 17-JUL-2002  
 LOCUS Danio rerio clone DKEY-224J12, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 10  
 DEFINITION unordered pieces.  
 ACCESSION AL840638  
 VERSION AL840638.3 GI:21911823  
 KEYWORDS HTG; HTGS\_PHASE1.  
 SOURCE Danio rerio.  
 ORGANISM Danio rerio



```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
REFERENCE
AUTHORS
1 (bases 1 to 188136)
TITLE
JOURNAL
DIRECT SUBMISSION
Submitted (14-JUL-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zface@sanger.ac.uk
Clone requests: clonerequest@sanger.ac.uk
On Jul 19, 2002 this sequence version replaced gi:21748404.
COMMENT
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zface@sanger.ac.uk
----- Project Information
Center project name: zk224j12
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 181961 bases at least Q40
Consensus quality: 184055 bases at least Q30
Consensus quality: 185651 bases at least Q20
Insert size: 187236; sum-of-contigs
Insert size: 157949; 24.5% error; agarose-fp
Quality coverage: 12.05x in Q20 bases; sum-of-contigs Quality
coverage: 14.71x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 4964: contig of 4964 bp in length
* 4965 5064: gap of 100 bp
* 5065 13025: contig of 7961 bp in length
* 13026 13125: gap of 100 bp
* 13126 64622: contig of 51497 bp in length
* 64623 64722: gap of 100 bp
* 64723 73250: contig of 8528 bp in length
* 73251 73350: gap of 100 bp
* 73351 125837: contig of 52487 bp in length
* 125838 125937: gap of 100 bp
* 125938 156004: contig of 30067 bp in length
* 156005 156104: gap of 100 bp
* 156105 159322: contig of 3218 bp in length
* 159323 159422: gap of 100 bp
* 159423 161843: contig of 2421 bp in length
* 161844 161943: gap of 100 bp
* 161944 165041: contig of 3098 bp in length
* 165042 165141: gap of 100 bp
* 165142 188136: contig of 22995 bp in length.
FEATURES
source
1..188136
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone="DKEX-224J12"
/clone_lib="Daniokeypilot"
1..4964
/note="assembly_fragment:00064
fragment_chain:1"
5065..13025
/note="assembly_fragment:02032
fragment_chain:1"
13126..64622
/note="assembly_fragment:00273
fragment_chain:1"
64723..73250
/note="assembly_fragment:02919
fragment_chain:1"
misc_feature
1..4964
/note="assembly_fragment:00407"
fragment_chain:1"
5065..13025
/note="assembly_fragment:00768"
fragment_chain:1"
13126..64622
/note="assembly_fragment:02758"
fragment_chain:1"
64723..73250
/note="assembly_fragment:02989"
fragment_chain:1"
165142..188136
/note="assembly_fragment:03788.0"
fragment_chain:1"
907 others

```

```

73351..125837
/note="assembly_fragment:01790
fragment_chain:1"
125838..156004
/note="assembly_fragment:00407"
fragment_chain:1"
156105..159322
/note="assembly_fragment:00768"
fragment_chain:1"
159423..161843
/note="assembly_fragment:02758"
fragment_chain:1"
161944..165041
/note="assembly_fragment:02989"
fragment_chain:1"
165142..188136
/note="assembly_fragment:03788.0"
fragment_chain:1"
907 others
BASE COUNT 60355 a 33818 c 33700 g 59356 t
ORIGIN
Alignment Scores:
Pred. No.: 846 Length: 188136
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.55% Indels: 0
DB: 2 Gaps: 0
US-09-966-880A-8 (1-198) x AL840638 (1-188136)
Qy 59 LeuLeuPheLeuArgTyrIleSerAsp 67
|||||
Db 35708 CTGCTTTTCCTCGCTTACATCTCAGAT 35734
|||||
RESULT 70
AC036188
LOCUS
DEFINITION Homo sapiens chromosome 11 clone RP11-287H12 map 11, WORKING DRAFT
SEQUENCE, 19 unordered pieces.
AC036188 190960 bp DNA linear HTG 24-AUG-2002
VERSION AC036188.2 GI:7770585
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 190960)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 11, clone RP11-287H12
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 190960)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bida,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karakas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrim,J., Meneus,D., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Olivari,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Triglio,J.,
Vassiliev,H., Viel,R., Vo.A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (07-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE
3 (bases 1 to 190960)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

```

```

Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castile,A., Choepel,Y., Colangelo,M., Collins,S.,
Collimore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Glnde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Laroque,K., Lamazares,R., Landers,T., Lehoczyk,J.,
Levine,J., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Olivat,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission
Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 12, 2000 this sequence version replaced gi:7523857.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L9252
Center clone name: 287_H_12
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 178632 bases at least Q40
Consensus quality: 185374 bases at least Q30
Consensus quality: 187834 bases at least Q20
Insert size: 194000; agarose-fp
Insert size: 189160; sum-of-contigs
Quality coverage: 4.2 in Q20 bases; agarose-fp
Quality coverage: 4.3 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 19 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1629: contig of 1629 bp in length
* 1630 1729: gap of 100 bp
* 1730 3289: contig of 1560 bp in length
* 3290 3389: gap of 100 bp
* 3390 4902: contig of 1513 bp in length
* 4903 5002: gap of 100 bp
* 5003 8987: contig of 3985 bp in length
* 8988 9087: gap of 100 bp
* 9088 12144: contig of 3057 bp in length
* 12145 12244: gap of 100 bp
* 12245 16220: contig of 3976 bp in length
* 16221 16320: gap of 100 bp
* 16321 20626: contig of 4306 bp in length
* 20627 20726: gap of 100 bp
* 20727 25379: contig of 4653 bp in length
* 25380 25479: gap of 100 bp
* 25480 32157: contig of 6678 bp in length
* 32158 32257: gap of 100 bp
*
* 32258 42184: contig of 9927 bp in length
* 42185 42284: gap of 100 bp
* 42285 48095: contig of 5811 bp in length
* 48096 48195: gap of 100 bp
* 48196 59553: contig of 11358 bp in length
* 59554 59653: gap of 100 bp
* 59654 70846: contig of 11193 bp in length
* 70847 70946: gap of 100 bp
* 70947 87757: contig of 18811 bp in length
* 87758 87857: gap of 100 bp
* 87858 103500: contig of 15643 bp in length
* 103501 103600: gap of 100 bp
* 103601 119609: contig of 16009 bp in length
* 119610 119709: gap of 100 bp
* 119710 138055: contig of 18346 bp in length
* 138056 138155: gap of 100 bp
* 138156 158360: contig of 20205 bp in length
* 158361 158460: gap of 100 bp
* 158461 190960: contig of 32500 bp in length.

FEATURES
            Location/Qualifiers
            1..190960
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="11"
            /map="11"
            /clone_lib="RPC1-11 Human Male BAC"
            1..1629
            /note="assembly_fragment"
            1730..3289
            /note="assembly_fragment"
            3390..4902
            /note="assembly_fragment"
            5003..8987
            /note="assembly_fragment"
            9088..12144
            /note="assembly_fragment"
            clone_end:T7
            vector_side:left"
            12245..16220
            /note="assembly_fragment"
            16321..20626
            /note="assembly_fragment"
            20727..25379
            /note="assembly_fragment"
            25480..32157
            /note="assembly_fragment"
            32258..42184
            /note="assembly_fragment"
            42285..48095
            /note="assembly_fragment"
            clone_end:SP6
            vector_side:right"
            48196..59553
            /note="assembly_fragment"
            59654..70846
            /note="assembly_fragment"
            70947..87757
            /note="assembly_fragment"
            87858..103500
            /note="assembly_fragment"
            103601..119609
            /note="assembly_fragment"
            119710..138055
            /note="assembly_fragment"
            138156..158360
            /note="assembly_fragment"
            158461..190960
            /note="assembly_fragment"

BASE COUNT 54008 a 39431 c 41327 g 54391 t 1803 others
ORIGIN
Alignment Scores:

```

Pred. No.:	857	Length:	190960
Score:	9.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	4.55%	Indels:	0
DB:	2	Gaps:	0
US-09-966-880A-8 (1-198) x AC036188 (1-190960)			
Qy	39 AlaThrSerPheSerLeuAspPheGly 47		
Db	68637 GCACACAGTTTCTCTAGATTTCGT 68663		
RESULT 71			
NCBI4D6			
LOCUS	NCBI4D6 199386 bp DNA linear PLN 18-JAN-2002		
DEFINITION	Neurospora crassa DNA linkage group II BAC contig B14D6.		
ACCESSION	AL356173		
VERSION	AL356173.2 GI:18375977		
KEYWORDS	Neurospora crassa.		
SOURCE	Neurospora crassa		
ORGANISM	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Sordariales; Sordariaceae; Neurospora.		
REFERENCE	1		
AUTHORS	Schulte,U., Align,V., Hoheisel,J., Brandt,P., Fartmann,B., Holland,R., Nyakatura,G., Mewes,H.W. and Mannhaupt,G.		
JOURNAL	Unpublished		
AUTHORS	2 (bases 1 to 199386)		
TITLE	German Neurospora genome,project.		
JOURNAL	Submitted (18-JAN-2002) MIPS, Institut fuer Bioinformatik, GSF-Forschungszentrum fuer Umwelt und Gesundheit, GmbH, Ingolstaedter Landstrasse 1, D-85764 Neuherberg, FRG, E-mail: G.Mannhaupt@gsf.de Project Coordinator: Ulrich Schulte, Institute of Biochemistry, Heinrich-Heine-University, D-40225 Duesseeldorf, E-mail: ulrich-schulte@uni-duesseeldorf.de		
COMMENT	On Jan 25, 2002 this sequence version replaced gi:7899304. Sequencing was performed by MWG Biotech AG, Ebersberg, Germany, http://www.mwgdna.com		
FEATURES	this contig is an assembly of BAC 14D6 from 1 to: 69793, cosmid 23D9 from 69794 to: 109492 rev, library pLORIST6Xh, strain 74-OR-23-1A, BAC 519 from 109493 to: 133555 rev, and BAC 1207 from 133556 to: 199386 BAC clones (strain OR74A) and cosmid clones are available at the Fungal Genetic Stock Center, http://www.fgsc.net; Information on performance of analysis and a more detailed annotation of this entry and other sequences of linkage groups II and V can be viewed at: http://mips.gsf.de/proj/neurospora.		
	source	Location/Qualifiers	
	misc_feature	1..199386	
		/organism="Neurospora crassa"	
rRNA	1028..1147	/db_xref="taxon:5141"	
		/chromosome="6"	
gene	1760..8212	/note="overlap to BAC B23D6"	
		/product="5S ribosomal RNA"	
gene	5283..7427,7508..7831,8041..8212)	/gene="B14D6.010"	
		complement(join(1760..2499,2596..3187,3256..5189,5283..7427,7508..7831,8041..8212))	
CDS	5283..7427,7508..7831,8041..8212)	/gene="B14D6.010"	
		complement(join(1760..2499,2596..3187,3256..5189,5283..7427,7508..7831,8041..8212))	
repeat_region	exon	/note="no similarity at C-terminus	
		similarity to kinesin-like protein Kif21b, Mus musculus, TREMBL:AF202893.1	
		Contains ATP/GTP-binding site motif A (P-loop) [QSGAGKSL];Kinesin motor domain signature and profile [SKMHFVDLAGE]	
		/codon_start=1	
repeat_region	exon	/product="related to kinesin-like protein"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	
repeat_region	exon	/gene="B14D6.010"	
		7880..7918	
		/note="39 bp poly(t) tandem repeat"	
		complement(8041..8212)	

```

/number=6
13243..15390
/gene="B14D6.020"
CDS
13243..15390
/gene="B14D6.020"
/notes="similarity to product:unknown, clone pp2464, homo sapiens, TREMBL:AF218002.1"
/codon_start=1
/product="Conserved hypothetical protein"
/protein_id="CAD21209.1"
/db_xref="GI:18375979"
/translation="MSLHVPPSAQKNVLDLDDLDYDNFDTADDVRPISTSKTPARKEQ
GLSTVDLGDSEQLKSIQARTMTQKTKTPELIDISDGFIDIDITRSTIRGERTS
EPVQRQLSPVSESNKSLRLOAHTPAISQSLARLGPSSRSRSTVEGESSPTD
ISGPNISAAOPTYHNDNDNDPPASPPRETREOKGKGPASTTDPVINIDDDDDDDP
FADSLPKVNPGRCPRAATTAANAADPISSSAPLPKPSPHRSRSGRSRSPLESDS
EDIGADPLSNEDDDFPDLSLRFANRLGNDLTGTPFKTSASITTTGATRTKPAKG
GTTKTAERAKERKALEKSAERKQERKQKAAEKQRAAAEAANKLRK
ETATPEMVDLPSSLPPTAKIQMEEFLLKIDYKEINTWTSPVDNVRWRRAVKSRYNE
ERHHYDPIPETIEKIIIVLPAAEFAKLAMAGBHNLEAHVLKVRHFNHQTIVL
TEGLKLLSSNRKNNDPFAVRSRLAEDESASTSSSRRTNKKNDPMTISEQSIDA
ALLRLQLLYSMQIOETTCLODTAHLQLFTQNVAVAPYKRHQEDLYLMSAGFCMDSGO
VRTALGTAEAYRMQLQEVARTAPAMGIANIYIPRVQQLVRALEEGGPGTLEDIRKVI
NKREVEGKRVKAVSKRLWKIIFTRGDEMSTEV"
13243..15390
/gene="B14D6.020"
/number=1
17216..18567
/gene="B14D6.030"
Join(17216..17318,17488..18161,18247..18445,18503..18567)
/gene="B14D6.030"
/notes="contains EST gb:BF072791"
/codon_start=1
/product="putative protein"
/protein_id="CAD21210.1"
/db_xref="GI:18375980"
/translation="MENLRPLLSLDIDRTIGSFDFLEAENNESLMRLVFRHOFOR
TMDQRDSFGSAYSCYEDKIELFLTYAFSSNETTALFLNAGFGYMRKDKMCKFDE
ISKVCKVGHPLTLPYIRINVRNLNQRHLAAWABQMELFLACNTDEVNNIRKI
YNTSDSNFRRIGRHADNRSMISVLSSAAALCEADELRMAASFWSRLSDEKS
SEMTAHSQALKEERQALDSITAGFTROLEORMHALESSIHIMALEHSORNSQLNM
ATAQHQIDLSITTTTRDAKTMKGLAFMGALFLPGTFMSAIFSMPTFEFKIDKLWYFAA
AVPLTLTITFG"
17216..17318
/gene="B14D6.030"
/number=1
17319..17487
/gene="B14D6.030"
/number=1
17409..17443
/notes="35 bp tgtgtgtgtgtgtgtgtgtgt tandem repeat"
17488..18161
/gene="B14D6.030"
/number=2
18162..18246
/gene="B14D6.030"
/number=2
18247..18445
/gene="B14D6.030"
/number=3
18446..18502
/gene="B14D6.030"
/number=3
18503..18567
/gene="B14D6.030"
/number=4
19258..19377
/product="5S ribosomal RNA"
20697..21695

```

Alignment Scores:  
Pred. No.: 890 Length: 19936  
Score: 9.00 Matches: 9

```

Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.55% Indels: 0
DB: 8 Gaps: 0

US-09-966-880A-8 (1-198) x NCB14D6 (1-199386)

Qy 174 ArgGlnLeuArgArgileLeuLeuPro 182
Db 171653 AGACAACCTGCGCGCATCTTCTGCGCA 171679

RESULT 72
AC103892/2
LOCUS 207584 bp DNA linear HTG 03-MAY-2002
DEFINITION Rattus norvegicus clone RP31-108E11 strain Brown Norway, WORKING
DRAFT SEQUENCE, 10 ordered pieces.
AC103892
AC103892.2 GI:20429784
HTG: HTGS_PHASE2; HTGS_DRAFT.
KEYWORDS Rattus norvegicus.
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 207584)
Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C.,
Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Larcic,P.,
Haghighi,P., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E., Laric,P.,
Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
Margulies,E.H., Masello,C., Maskeri,B., Mastrian,S.D.,
McCluskey,J.C., McDowell,J., Paguirigan,C., Pearson,R.,
Portnov,M.E., Prasad,A., Schueler,M.G., Stantripop,S., Thomas,J.W.,
Thomas,P.J., Touchman,J.W., Tsurgon,C., Vogt,J.L., Walker,M.A.,
Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 207584)
Direct Submission
Green,E.D.
Submitted (30-NOV-2001) NIH Intramural Sequencing Center, 8717
Grovmont Circle, Gaithersburg, MD 20877, USA
3 (bases 1 to 207584)
Direct Submission
Green,E.D.
Submitted (03-MAY-2002) NIH Intramural Sequencing Center, 8717
Grovmont Circle, Gaithersburg, MD 20877, USA
On May 3, 2002 this sequence version replaced gi:17155044.
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_zoo@nhgri.nih.gov
----- Project Information
Center project name: ckc
Center clone name: 108E11

```

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8X average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics  
Sequencing vector: plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990319

Consensus quality: 205765 bases at least Q40  
 Consensus quality: 206455 bases at least Q30  
 Consensus quality: 206619 bases at least Q20  
 Insert size: 199000; agarose-fp  
 Quality coverage: 10.48x in Q20 bases; agarose-fp  
 Quality coverage: 10.09x in Q20 bases; sum-of-contigs  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 10 contigs. Gaps between the contigs  
 \* are represented as runs of N. The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* provided by the submittor.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.

1 9523: contig of 9523 bp in length  
 \* 9524 9623: gap of unknown length  
 \* 9624 19034: contig of 9411 bp in length  
 \* 19035 19134: gap of unknown length  
 \* 19135 28903: contig of 9769 bp in length  
 \* 28904 29003: gap of unknown length  
 \* 29004 54011: contig of 25008 bp in length  
 \* 54012 54111: gap of unknown length  
 \* 54112 69522: contig of 15411 bp in length  
 \* 69523 69622: gap of unknown length  
 \* 69623 135448: contig of 65826 bp in length  
 \* 135449 135548: gap of unknown length  
 \* 135549 144120: contig of 8572 bp in length  
 \* 144121 144220: gap of unknown length  
 \* 144221 171771: contig of 27551 bp in length  
 \* 171772 171871: gap of unknown length  
 \* 171872 199518: contig of 27647 bp in length  
 \* 199519 199618: gap of unknown length  
 \* 199619 207584: contig of 7966 bp in length.

FEATURES

source  
 1. .207584  
 /organism="Rattus norvegicus"  
 /strain="Brown Norway"  
 /db\_xref="taxon:10116"  
 /clone="RP31-108E11"  
 /clone\_lib="RP31"  
 1. .41024  
 /note="clone overlaps with GenBank Accession Number  
 AC097001 clone RP31-358M14 (center project name ckb)"  
 misc\_feature  
 1. .9523  
 /note="assembly\_fragment  
 clone\_end:17  
 vector\_side:left"  
 9624. .19034  
 /note="assembly\_fragment"  
 19135. .28903  
 /note="assembly\_fragment"  
 29004. .54011  
 /note="assembly\_fragment"  
 54112. .69522  
 /note="assembly\_fragment"  
 69623. .135448  
 /note="assembly\_fragment"  
 135449. .144120  
 /note="assembly\_fragment"  
 144221. .171771  
 /note="assembly\_fragment"  
 171872. .199518  
 /note="assembly\_fragment"  
 199619. .207584  
 /note="assembly\_fragment  
 clone\_end:SP6  
 vector\_side:right"  
 BASE COUNT 60061 a 46035 c 45902 g 54686 t 900 others  
 ORIGIN

Alignment Scores:  
 Pred. No.: 921 Length: 207584  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 4.55% Indels: 0  
 DB: 2 Gaps: 0  
 US-09-966-880A-8 (1-198) x AC103892 (1-207584)  
 QY 42 PheSerLeuAspPheGlyTyrLeuArg 50  
 DB 181210 TTTTCTTTGGATTTGGTTATTACGG 181184  
 RESULT 73  
 AC091322  
 LOCUS Mus musculus clone RP23-60D3, WORKING DRAFT SEQUENCE, 5 unordered  
 DEFINITION pieces.  
 AC091322  
 VERSION AC091322.2 GI:14336533  
 KEYWORDS HTG: HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
 SOURCE Mus musculus.  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 232951)  
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
 Mus musculus, clone RP23-60D3  
 Unpublished  
 2 (bases 1 to 232951)  
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S.,  
 Barna, N., Bastien, V., Boguslavsky, L., Bouckgalter, B., Brown, A.,  
 Camarata, J., Campopiano, A., Chang, J., Choepel, Y., Collangelo, M.,  
 Collins, S., Collymore, A., Cooke, P., DeArellano, K., Dewar, K.,  
 Diaz, J. S., Dodge, S., Fero, S., Ferreira, P., Fitzhugh, W., Gage, D.,  
 Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,  
 Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W.,  
 Iliev, I., Johnson, R., Jones, C., Karatas, A., LaRocque, K.,  
 Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,  
 MacLean, C., Macdonald, P., Marquis, N., Matthews, C., McCarthy, M.,  
 McEwan, P., McKernan, K., McPheters, R., Meldrum, J., Meneus, L.,  
 Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C.,  
 Norman, C. H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J.,  
 Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,  
 Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,  
 Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S.,  
 Severy, P., Sougne, C., Spencer, B., Stange-Thomann, N.,  
 Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S.,  
 Theodore, J., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,  
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,  
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (15-APR-2001) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Jun 11, 2001 this sequence version replaced gi:13625486.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence.submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: LI3252  
 Center clone name: 60\_D3  
 ----- Summary Statistics  
 Sequencing vector: plasmid; n/a; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 231176 bases at least Q40  
 Consensus quality: 231965 bases at least Q30

Consensus quality: 232319 bases at least Q20  
 Insert size: 210000; agarose-fp  
 Insert size: 232551; sum-of-contigs  
 Quality coverage: 10.7 in Q20 bases; agarose-fp  
 Quality coverage: 9.6 in Q20 bases; sum-of-contigs  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 5 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \*  
 \* 1 53: contig of 53 bp in length  
 \* 154 153: gap of 100 bp  
 \* 154 2590: contig of 2437 bp in length  
 \* 2591 2690: gap of 100 bp  
 \* 2691 11774: contig of 9084 bp in length  
 \* 11775 11874: gap of 100 bp  
 \* 11875 100524: contig of 88650 bp in length  
 \* 100525 100624: gap of 100 bp  
 \* 100625 232951: contig of 132327 bp in length.  
 \* Location/Qualifiers  
 \* 1..232951  
 \* /organism="Mus musculus"  
 \* /db\_xref="taxon:10090"  
 \* /clone="RP23-60D3"  
 \* /clone\_lib="RPCI-23 Female Mouse BAC"  
 \* 1..53  
 \* /note="assembly\_fragment"  
 \* clone\_end:sp6  
 \* vector\_side:left  
 \* 154..2590  
 \* /note="assembly\_fragment"  
 \* 2691..11774  
 \* /note="assembly\_fragment"  
 \* 11875..100524  
 \* /note="assembly\_fragment"  
 \* 100625..232951  
 \* /note="assembly\_fragment"  
 \* clone\_end:t7  
 \* vector\_side:right"  
 \* 70795 a 45474 g 46842 g 69434 t 406 others  
 \* ORIGIN  
 \*  
 \* Alignment Scores:  
 \* Pred. No.: 1.02e+03 Length: 232951  
 \* Score: 9.00 Matches: 9  
 \* Percent Similarity: 100.00% Conservative: 0  
 \* Best Local Similarity: 100.00% Mismatches: 0  
 \* Query Match: 4.55% Indels: 0  
 \* DB: 2 Gaps: 0  
 \*  
 \* US-09-966-880A-8 (1-198) x AC091322 (1-232951)  
 \* Qy 20 TrpAlaLySGlyArgArgGluThrTyr 28  
 \* |||||  
 \* Db 143785 TGGCAAAAGGCGAGGAGACTTAT 143811  
 \*  
 \* RESULT 74  
 \* AE003632  
 \* LOCUS  
 \* DEFINITION Drosophila melanogaster genomic scaffold 142000013386055 section 25  
 \* of 63, complete sequence.  
 \* ACCESSION AE003632 AE002690  
 \* VERSION AE003632.2 GI:10728720  
 \* KEYWORDS HTG.  
 \* SOURCE  
 \* ORGANISM Drosophila melanogaster.  
 \* Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;  
 \* Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 \* Ephydroidea; Drosophilidae; Drosophila.

REFERENCE  
AUTHORS

1 (bases 1 to 244237)  
 Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,  
 Ananides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Gallo,R.F.,  
 George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,  
 Sutton,G.C., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,  
 Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,  
 Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor  
 Miklos,G.L., Abril,J.F., Aghayani,A., An,H.J.,  
 Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,  
 Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,  
 Benos,P.V., Bereman,B.P., Bhandari,D., Bolshakov,S., Borkov,D.,  
 Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Hurlis,K.C.,  
 Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,  
 Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de  
 Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,  
 Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., DunKov,R.C.,  
 Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferrier,S.,  
 Fleischmann,W., Fowler,C., Gabriellian,A.E., Garg,N.S.,  
 Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,  
 Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,  
 Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,  
 Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,  
 Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,  
 Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,  
 Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattel,B., McIntosh,T.C.,  
 McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,  
 Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,H., Murphy,L.,  
 Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,  
 Nusskern,D.R., Pacieb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,  
 Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,  
 Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Sidon-Kiamos,I.,  
 Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,  
 Stapleton,M., Strong,R., Sun,E., Svirskaas,R., Tector,C., Turner,R.,  
 Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wasserman,D.A.,  
 Weinstein,G.M., Weissbach,J., Williams,S.M., Woodage,T.,  
 Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,  
 Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,  
 Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,  
 Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.  
 The genome sequence of *Drosophila melanogaster*  
 Science 287 (5461), 2185-2195 (2000)  
 20196006  
 10731132  
 2 (bases 1 to 244237)  
 Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.  
 Direct Submission  
 TITLE  
 Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,  
 Rockville, MD, USA  
 COMMENT  
 On Oct 9, 2000 this sequence version replaced gi:7297846.  
 FEATURES  
 source  
 1..244237  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 /chromosome="2L"  
 complement(<28430..>29119)  
 /gene="CG14926"  
 /evidence=not\_experimental  
 /map="32F2-32F2"  
 /db\_xref="FLYBASE:FBan0014926"  
 /db\_xref="FLYBASE:FBgn0032360"  
 complement(join(<28430..28479,28534..28730,28788..28996,  
 29051..>29119))  
 /gene="CG14926"  
 /product="CT34753"  
 /evidence=not\_experimental  
 /db\_xref="FLYBASE:FBan0014926"  
 /db\_xref="FLYBASE:FBgn0032360"  
 complement(join(28430..28479,28534..28730,28788..28996,  
 29051..29119))  
 /gene="CG14926"  
 /note="CG14926 gene product"  
 /codon\_start=1  
 /evidence=not\_experimental

```
/protein_id="AAF53095.1"
/db_xref="GI:7297847"
/db_xref="FLYBASE:FBan0014926"
/db_xref="FLYBASE:FBan0032360"
/translation="MENGPEWETSHGQGNPNYSQMGSTPPGGGROWLHSSQA
NTQYLHGSSGMSGHNNASPDGFASYNNQNMNYTNFKPSYSHSQVGPASVQMGMS
NDPNPSQSHRGSWF"
<52364..>59160
/gene="salr"
/note="CG4881"
/map="32f2-32f3"
/db_xref="FLYBASE:FBan0004881"
/db_xref="FLYBASE:FBan0000287"
join(<52364..54959,55677..55837,57830..58091,58154..58386,
58476..58673,58807..>59160)
/gene="salr"
/product="Ct15643"
/note="Nucleotide sequence of the Celera sequence differs
from the published sequence for this transcript."
/db_xref="FLYBASE:FBan0004881"
/db_xref="FLYBASE:FBan0000287"
join(52364..54959,55677..55837,57830..58091,58154..58386,
58476..58673,58807..59160)
/gene="salr"
/note="salr gene product; Nucleotide sequence of the
Celera sequence differs from the published sequence for
this transcript"
/codon_start=1
/protein_id="AAF53096.1"
/db_xref="GI:7297848"
/db_xref="FLYBASE:FBan0004881"
/db_xref="FLYBASE:FBan0000287"
/translation="MCSIFRNRIYRCGTGTSRSGERDRDRDRDRDRDRDRT
LLAELDADSNNGTEPOMEAEVPESTRETAEREEPEEPENSALDLSISSG
GRESLPGSGHVSLEALQHTKVAAQFAATAAGNHQSADLAWQSTIFNVQRHLMQL
QLIQHQLRAAALAGRHSHSDEEEPEPEPKKQPTNGLKEELEQGPSE
QDSREENREKTRGTEDEKRAPEGYQSMWCIDISSLASSIITNHDPAPNPEPC
LEMJQRTEEVDSASLSHAAQOEYSEYSAEQSRGEIFKHKYCKCKIFGSYS
ALQJLHRSHTGERPFCVNCGSKTTKGNLKVHVORHTQIFPMLLPGVAPNVGHS
QGQVGEYPIRLFPFAPVAVPGQEHQONQVEEIEIKQEIIPQAEDELSPMKKEE
KSHSVKRTPKVKTDALPSSKEPKESIKPVYTSRRNGSVRRKQTSVAPPPQE
DREDLVHLHLIAKLRSSASRESQPAEYSLAQMERTIGDSWEDLIEDTSETSL
QQLVDNIENLTQNOCLFCQVMSRCSLQMHIRTHTGKPFCKICGRAPAKNPG
KAHMSIKIKPMRSQKPCVYCHOKFSNGILQOHIRHTMDGSGGGAAPANPEA
ERLGTEDONSNGISLGTDLDFSTLSDHSGORSESSOGGDFDEMTWSDSDRNS
SAATATPHLERDRERERRIPNDCSDRSRNSPDLTGRESGEMPAWDLSSPS
SNSGRIFATLANGAAGGSGNGGLPMLGMPMPNLLMAAAREEMHALGHAHAKFPL
LPFGPLGPMGLHPPNVNCLCFKMLPSLALESLOSEHAKEPATGHAQRQPSDAGS
PYGAKLTUNPLNLFARKPPSSSSSSSEKLPSSNPPFPAENPPATPIKEDPDQOLMV
EEGASGSGTGATSNYPQAGDAEQSLMKWLHAHFRFPASPLDFOCALMSAGPPTS
SLDPVNNKHFCVCRNRFSSSALOIHMTHTGDKPQCNCVOKAFYTKGNLKVHMG
THMTNTPSRGRGNLRLMELPFGNCSQGHGSAEQEFMQRPELFFLYLPFPFNG
LPPKPGSLGAFNPPIPPPPFANGKYPYPPGLLFGPGFLAHPYILERRSSKSPTP
EPAQSPALREEGSGNIWHPLSRKVENNQNESMGGFNEQEDHAETDGTGDNDESES
RDABK"
complement(<122109..>131129)
/gene="salr"
/note="CG464"
/map="33A1-33A1"
/db_xref="FLYBASE:FBan0006464"
/db_xref="FLYBASE:FBan0004579"
complement(join(<122109..122243,122367..126202,
130979..>131129))
/gene="salr"
/product="CT20082"
/note="Nucleotide sequence of the Celera sequence differs
from the published sequence for this transcript."
/db_xref="FLYBASE:FBan0006464"
/db_xref="FLYBASE:FBan0004579"
complement(join(122109..122243,122367..126202,
130979..131129))
/gene="salr"
```

```

/note="salm gene product; Nucleotide sequence of the
Celera sequence differs from the published sequence for
this transcript"
/codon_start=1
/protein_id="AAF53097.1"
/db_xref="GI:7297849"
/db_xref="FLYBASE:FBan0006464"
/db_xref="FLYBASE:FBan0004579"
/translation="MKNHLSNVLCAMRSDFKDNHOETINKMIOFTGVYGIKOLKDR
ARSADKIGSDQEBEGGCSPLTATTATSPRSPEPEPEEESQSPPEQSTPD
HOLENDIKSEAKSEIEPVEDNNNRVMTKPSSEERENASMPSSPVAESAEEAAT
ERTPEKEKEDVDEVEKDPAPSSAVPSTVTLPGGAGAPVTLEAIONQMAIAQFA
AKTITANGSKADNEAAMKQLAFLOQTFLNQOQLFOQLIQQLSQSLALNAQOEED
TEEDAOEODEOQETDTVEEBERADLMELROKAEARMAEAKARQHLINAGVPLRESSG
SPASLKRPRRHDSOPNRPISLNDTHKADTAODALAKLAKEMENTPLPFGSGLASSI
ITNDDULPEPNSLOLLQKRAQEVLDLSAQGLANSMDDFAFGEKSGEGKGRNPFKK
HRCYCGKVGSGSALQIIRKSHSTGFRFCNVCGRSTTKGNLKVHFRQHAQKPPHY
PMNATPIPEHMDKHFPLLDQMSPTSDSPNSPAPPLGSPAPFPAPGLQNLRYR
MEILKSLGAAAPHQYPOELPTDLRKPSQOLDEPOVKNPEVVEEKDOREHEQEMA
ECSEPEPEPLPLEVRIKEERVEEQKOEHDRIEPRTPSPSEHRSPPHRRSHMG
YPPVQPTOPAAALMHPOSSQSHLOHLPQOLPPREDFEAFERPLNFTTAKMLSP
EHSFSPRAGCALPPGVPPPHHHHMAKSPFNPIKHEMAALLPRPHSNDNSWEN
FIEVNTCETWKLKELMKNKISIDPNCVCDRLSCSALOMHYRTHTGERPFCKRI
CGRAFTKGNLTKTHMAVHKIRPMRNFHQPCVHKYSNALVLQOHLRLHGTGEPTDLT
PEQIAAEIRDPPSPMPPGHFMPNFAAAAFHFGALPGGPGPPGPHGAHNGALGES
SOGMDDNMGCEGYDDDDVSEHLSNSNLEQDGRSGDDFKSLLPQKLRIDATGV
VNTNVPRESSASSHGHVSGTSAPTSPVHASSQVTKRSSPARSEASQALDITPR
AAPTSSSSRSPLEKRPVPPSPRSPGSHASANILTSPLPTVGDICLPPGLQH
HLQOOHQLMQOQAASAAAAAHHHOOAALHOOHQLRREAAEOOAKAAAAAA
AAAAAQRTTPQARDQEGGAGAGPPNPMGAPFPMPLPFPATQNMCN
AMNQIAQSVMPAAPPNPLALSGVRGTTGICYKTFPCGSALEHYRSHTERPFKCS
ICDRGFTKGNLKHQMLTHKTRDMEQETFRNRAVKYVSLPISPIPSYPFHSYS"
<172159..>172896
/gene="sala"
/note="CG4922"
/map="33A1-33A2"
/db_xref="FLYBASE:FBan0004922"
/db_xref="FLYBASE:FBan0003313"
join(172159..172172,172243..172896)
/gene="sala"
/product="Ct15718"
/note="Nucleotide sequence of the Celera sequence differs
from the published sequence for this transcript."
```

Alignment Scores:	Pred. No.:	Length:	244237
Score:	1.06e+03	Matches:	9
Percent Similarity:	9.00	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	4.55%	Indels:	0
DB:	3	Gaps:	0

US-09-966-880A-8 (1-198) x AE003632 (1-244237)

QY	117	GluAspArgLysAlaGluProGluGly	125
Db	53072	GAGGACAGAAAGGACAGACGAGCGGT	53098

RESULT 75	TCARITS1	245 bp	DNA	linear	PLN 03-AUG-1999
LOCUS	TCARITS1				
DEFINITION	Troutvetteria carolinensis internal transcribed spacer 1, complete sequence.				
ACCESSION	U96035				
VERSION	U96035.1	GI:4205107			
KEYWORDS					
SEGMENT					
SOURCE					
ORGANISM					
REFERENCE					

1 of 2  
Troutvetteria carolinensis.  
Troutvetteria carolinensis  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; Ranunculales;  
Ranunculaceae; Troutvetteria.  
1 (bases 1 to 245)

AUTHORS Xiang,Q.Y., Soltis,D.E. and Soltis,P.S.  
 TITLE The eastern Asian and western North American floristic  
 disjunction: congruent phylogenetic patterns in seven diverse  
 genera

JOURNAL Mol. Phylogenet. Evol. 10 (2), 178-190 (1998)  
 MEDLINE 99097283  
 PUBMED 9878229

REFERENCE 2 (bases 1 to 245)  
 AUTHORS Xiang,Q.-Y., Soltis,D.E. and Soltis,P.S.  
 TITLE Direct Submission  
 JOURNAL Submitted (02-APR-1997) Plant Biology, Ohio State University, 1735  
 Neil Avenue, Columbus, OH 43210, USA

FEATURES  
 source  
 1..245  
 Location/Qualifiers  
 /organism="Trautvetteria carolinensis"  
 /db\_xref="taxon:46993"  
 misc\_RNA  
 1..245  
 /product="internal transcribed spacer 1"  
 /note="ITS1"

BASE COUNT 65 a 65 c 64 g 50 t 1 others  
 ORIGIN

Alignment Scores:  
 Pred. No.: 31.3 Length: 245  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 4.04% Indels: 0  
 DB: 0

US-09-966-880A-8 (1-198) x TCARITS1 (1-245)

Qy 174 ArgGlnLeuArgArgIleLeuLeu 181  
 Db 173 GCCCAATTCGCCGATTTTGTG 150

RESULT 76  
 HSA279528  
 LOCUS HSA279528 354 bp mRNA linear PRI 01-JUN-2000  
 DEFINITION Homo sapiens mRNA for immunoglobulin mu heavy chain variable  
 region, partial, clone CD27high-9.

ACCESSION AJ279528  
 VERSION AJ279528.1 GI:6723553  
 KEYWORDS IgM; IgM heavy chain; immunoglobulin mu heavy chain; variable  
 region.

SOURCE  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 354)  
 Dono,M., Zupo,S., Leanza,N., Mellioli,G., Fogli,M., Melagrana,A.,  
 Chiorazzi,N. and Ferrarini,M.  
 TITLE Heterogeneity of tonsillar subepithelial B lymphocytes, the splenic  
 marginal zone equivalents  
 JOURNAL J. Immunol. 164 (11), 5596-5604 (2000)  
 MEDLINE 20281644  
 PUBMED 10820234

REFERENCE 2 (bases 1 to 354)  
 AUTHORS Dono,M.  
 TITLE Direct Submission  
 JOURNAL Submitted (11-JAN-2000) Dono M., Istituto Nazionale per la Ricerca  
 sul Cancro, Immunologia Clinica, L.go Rosanna benzi 10- Genova,  
 ITALY

FEATURES  
 source  
 1..354  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="14"  
 /clone="CD27high-9"  
 /cell\_type="IgD-low SE B-lymphocyte"  
 /tissue\_type="tonsil"  
 /rearranged  
 1..354

gene

CDS  
 /gene="IGHV4-34"  
 <1..>354  
 /gene="IGHV4-34"  
 /codon\_start=1  
 /product="Immunoglobulin mu heavy chain variable region"  
 /protein\_id="CAB66358.1"  
 /db\_xref="GI:6723554"  
 /translation="QVLLQMGAGLLKPKSETLSLTCAVYGGSFNDYYWSMIRQPPGKG  
 LEWIGNLYYSGSTYNSLSKSRVTIFVDTSKNQFSLKLSSTAAATAVYYCARHVA  
 RRDYWGQGTLLTVSS"  
 <1..>354  
 V\_region  
 /gene="IGHV4-34"  
 /product="Immunoglobulin mu heavy chain variable region"  
 74 a 99 c 98 g 83 t

BASE COUNT  
 ORIGIN

Alignment Scores:  
 Pred. No.: 43.1 Length: 354  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 4.04% Indels: 0  
 DB: 0

US-09-966-880A-8 (1-198) x HSA279528 (1-354)

Qy 90 CysAlaArgHisValAlaAspPhe 97  
 Db 283 TGTGCGAGACACGTCGCGGACTTT 306

RESULT 77  
 AX389702/C  
 LOCUS AX389702 500 bp DNA linear PAT 19-MAR-2002  
 DEFINITION Sequence 4630 from Patent WO214500.  
 ACCESSION AX389702  
 VERSION AX389702.1 GI:19582832  
 KEYWORDS human.  
 SOURCE

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1  
 Escobedo,J., Garcia,P.D., Sudduth-Klinger,J., Reinhard,C.,  
 Randazzo,F., Lamson,G., Scott,E.M., Zhang,G., Kassam,A., Pot,D. and  
 Labat,I.  
 TITLE Human genes and gene expression products  
 JOURNAL Patent: WO 0214500-A 4630 21-FEB-2002;  
 CHIRON CORPORATION (US) ; Hyseq Inc. (US)  
 FEATURES  
 source  
 1..500  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"

BASE COUNT 227 a 75 c 72 g 126 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 58.1 Length: 500  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 4.04% Indels: 0  
 DB: 6

US-09-966-880A-8 (1-198) x AX389702 (1-500)

Qy 59 LeuLeuPheLeuArgTyrIleSer 66  
 Db 264 TTGTTGTTTCTAAGATACATATCT 241

RESULT 78  
 HSC07H12  
 LOCUS HSC07H12 569 bp DNA linear STS 01-MAR-2000  
 DEFINITION H.sapiens STS from flow-sorted chromosome X random shear fragment,



```

sequence tagged site.
ACCESSION AL158650
VERSION AL158650.1 GI:7161252
KEYWORDS STS.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 569)
AUTHORS Hunt,S., Pavitt,B., Willey,D., Carter,N. and Ross,M.
TITLE Direct Submission
JOURNAL Submitted (02-MAR-2000) E-mail contact: humquerry@sanger.ac.uk
COMMENT Marker stFSX1703 (Primer A : CTCCTCAGAAGCATTCCTG; Primer B :
TGGTTTATGCTTGTGTTTCCA; amplicon size : 184 bp) was developed from
a single pass sequencing read from H.sapiens flow-sorted chromosome
X random shear fragment, SCXpc07h12. Vector : pUC18 Site : SmaI
Further information : http://www.sanger.ac.uk/HGP/ChrX/.
FEATURES
    Location/Qualifiers
        1..569
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="X"
            /clone="SCXpc07h12"
            /sex="male"
            /tissue_type="Lymphoblastoid cell line"
            /clone_lib="SCXpc"
BASE COUNT 203 a 92 c 76 g 198 t
ORIGIN
Alignment Scores:
Pred. No.: 64.9 Length: 569
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.04% Indels: 0
DB: 11 Gaps: 0

US-09-966-880A-8 (1-198) x HSC07H12 (1-569)
Qy 138 lleMetThrPheLysAspTyrPhe 145
Db 306 ATCATGACATTAAAGATTATTTT 329

RESULT 79
LOCUS PM3C3B 653 bp DNA linear STS 09-MAR-2002
DEFINITION Penicillium marneffei STS, clone pm3c3.b, sequence tagged site.
ACCESSION AL685634
VERSION AL685634.1 GI:19336682
KEYWORDS STS.
SOURCE Penicillium marneffei.
ORGANISM Penicillium marneffei
REFERENCE 1
AUTHORS Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.
Yuen,K.Y., Pascal,G., Wong,S., Glaser,P., Woo,P., Kunst,P.,
Cheung,E., Medigue,C. and Danchin,A.
TITLE Exploring the Penicillium marneffei genome
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 653)
AUTHORS Danchin,A. and Pascal,G.
TITLE Direct Submission
JOURNAL Submitted (08-MAR-2002) Danchin A., HKU-Pasteur Research Centre,
Dexter HC Man Building 8, Sassoon Road, Pokfulam, Hong Kong
FEATURES
    Location/Qualifiers
        1..653
            /organism="Penicillium marneffei"
            /db_xref="taxon:37727"
            /clone="pm3c3.b"
BASE COUNT 176 a 159 c 140 g 178 t
ORIGIN

Alignment Scores:
Pred. No.: 176 a 159 c 140 g 178 t
Score: 104 Length: 653
Percent Similarity: 100.00% Conservative: 8
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.04% Indels: 0
DB: 11 Gaps: 0

US-09-966-880A-8 (1-198) x A48781 (1-984)
Qy 176 LeuArgArgIleLeuLeuProLeu 183
Db 837 CTCCGACGATCCTTCTTCCTCTA 860

RESULT 81
LOCUS A58739 984 bp DNA linear PAT 06-MAR-1998
DEFINITION Sequence 2 from Patent WO9700319.
ACCESSION A58739
VERSION A58739.1 GI:3714281
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 984)
AUTHORS Browne,M.J., Chapman,C.G., Clinkenbeard, Helen,E. and Robinson,J.H.
TITLE CHIMERIC LEPTIN FUSED TO IMMUNOGLOBULIN DOMAIN AND USE
JOURNAL Patent: WO 9700319-A 2 03-JAN-1997;
COMMENT Other publication AU 6011096 970115.
FEATURES
    Location/Qualifiers
        1..984
            /organism="unidentified"
            /db_xref="taxon:32644"
BASE COUNT 230 a 322 c 269 g 163 t
ORIGIN

Alignment Scores:
Pred. No.: 104 Length: 984
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.04% Indels: 0
DB: 6 Gaps: 0

US-09-966-880A-8 (1-198) x PM3C3B (1-653)
Qy 171 ArgLeuSerArgGlnLeuArgArg 178
Db 447 CGGCTCAGTCGCCAACAACGTACGTCGT 470

RESULT 80
LOCUS A48781 984 bp DNA linear PAT 07-MAR-1997
DEFINITION Sequence 8 from Patent WO9604388.
ACCESSION A48781
VERSION A48781.1 GI:2302449
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 984)
AUTHORS Browne,M.J., Murphy,K.E., Chapman,C.G., Clinkenbeard,H.E.,
Young,P.R. and Shatzman,A.R.
TITLE NOVEL COMPOUNDS
JOURNAL Patent: WO 9604388-A 8 15-FEB-1996;
COMMENT SMITHKLINE BEECHAM PLC (GB)
Other publication AU 3382595 960304.
FEATURES
    Location/Qualifiers
        1..984
            /organism="unidentified"
            /db_xref="taxon:32644"
BASE COUNT 230 a 322 c 269 g 163 t
ORIGIN

```

```
Alignment Scores: 104 Length: 984
Pred. No.: 8.00 Matches: 8
Score: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.04% Indels: 0
DB: 6 Gaps: 0

US-09-966-880A-8 (1-198) x A58739 (1-984)

Qy 176 LeuArgArgIleLeuLeuProLeu 183
Db 837 CTCGACGCGATCCTTCTTCCTCTA 860

RESULT 82
AR018707
LOCUS AR018707 984 bp DNA linear PAT 05-DEC-1998
DEFINITION Sequence 8 from patent US 5783181.
ACCESSION AR018707
VERSION AR018707.1 GI:3973821
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 984)
AUTHORS Browne,M.Joseph., Young,P.Ronald., Shatzman,A.Richard.,
Murphy,K.Elizabeth., Chapman,C.Gerald. and
Clinkenbeard,H.Elizabeth.
TITLE Therapeutic uses of fusion proteins between mutant IL 4/IL13
JOURNAL antagonists and immunoglobulins
FEATURES Patent: US 5783181-A 8 21-JUL-1998;
Location/Qualifiers
source 1..984
/organism="unknown"
BASE COUNT 230 a 322 c 269 g 163 t
ORIGIN
Alignment Scores: 104 Length: 984
Pred. No.: 8.00 Matches: 8
Score: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.04% Indels: 0
DB: 6 Gaps: 0

US-09-966-880A-8 (1-198) x AR018707 (1-984)

Qy 176 LeuArgArgIleLeuLeuProLeu 183
Db 837 CTCGACGCGATCCTTCTTCCTCTA 860

RESULT 83
A48778
LOCUS A48778 1006 bp DNA linear PAT 07-MAR-1997
DEFINITION Sequence 5 from patent WO9604388.
ACCESSION A48778
VERSION A48778.1 GI:2302447
KEYWORDS unidentified.
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1006)
AUTHORS Browne,M.J., Murphy,K.E., Chapman,C.G., Clinkenbeard,H.E.,
Young,P.R. and Shatzman,A.R.
TITLE NOVEL COMPOUNDS
JOURNAL Patent: WO 9604388-A 5 15-FEB-1996;
SMITHKLINE BEECHAM PLC (GB)
COMMENT Other publication AU 3382595 960304.
FEATURES Location/Qualifiers
source 1..1006
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 235 a 327 c 271 g 173 t
ORIGIN
Alignment Scores: 106 Length: 1006
Pred. No.: 8.00 Matches: 8
Score: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.04% Indels: 0
DB: 6 Gaps: 0

US-09-966-880A-8 (1-198) x AR018705 (1-1006)

Qy 176 LeuArgArgIleLeuLeuProLeu 183
Db 837 CTCGACGCGATCCTTCTTCCTCTA 860

RESULT 85
AY094934
LOCUS AY094934 1101 bp mRNA linear INV 15-APR-2002
DEFINITION Drosophila melanogaster RH52423 full insert cDNA.
ACCESSION AY094934
VERSION AY094934.1 GI:20151854
KEYWORDS FLI_CDNA.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Carlson,J.,
Champe,M., Chavez,C., Dorsett,V., Dresnek,D., Farfan,D., Frise,E.,
George,R., Gonzalez,M., Guarin,H., Krommiller,B., Li,P., Liao,G.,
Miranda,A., Mungall,C.J., Nunoo,J., Pacleb,J., Parauas,V., Park,S.,
Patel,S., Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Rubin,G.M.
and Celinker,S.
```

TITLE Direct Submission  
JOURNAL Submitted (03-APR-2002) Berkeley Drosophila Genome Project, Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA 94720, USA

COMMENT Sequence submitted by:  
Berkeley Drosophila Genome Project  
Lawrence Berkeley National Laboratory  
Berkeley, CA 94720

This clone was sequenced as part of a high-throughput process to sequence clones from Drosophila Gene Collection 1 (Rubin et al., Science 2000). The sequence has been subjected to integrity checks for sequence accuracy, presence of a polyA tail and contiguity within 100 kb in the genome. Thus we believe the sequence to reflect accurately this particular cDNA clone. However, there are artifacts associated with the generation of cDNA clones that may have not been detected in our initial analyses such as internal priming, priming from contaminating genomic DNA, retained introns due to reverse transcription of unspliced precursor RNAs, and reverse transcriptase errors that result in single base changes. For further information about this sequence, including its location and relationship to other sequences, please visit our web site (<http://fruitfly.berkeley.edu>) or send email to [cdna@fruitfly.berkeley.edu](mailto:cdna@fruitfly.berkeley.edu).

FEATURES  
Source Location/Qualifiers  
1..1101  
/organism="Drosophila melanogaster"  
/strain="y: cn bw sp"  
/db\_xref="taxon:7227"  
misc\_feature 1..1091  
/note="sim4 alignment with AE003477 (63B6-63D1)"  
CDS 750..1010  
/note="Longest ORF"  
/codon\_start=1  
/product="RH52423p"  
/protein\_id="AAM11287.1"  
/db\_xref="GI:20151856"  
/translation="MFNTKMYSVRIWLGSVLLGPLLLAAIAVQGOEPASPVFQNH  
RTKEWNLNDNTFSDCKRRSVGFYADMEYNCQIFHMCDEY"  
BASE COUNT 343 a 213 c 240 g 305 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 115 Length: 1101  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 4.04% Indels: 0  
DB: 3 Gaps: 0  
US-09-966-880A-8 (1-198) x AY094934 (1-1101)  
QY 42 PheSerLeuAspPheGlyTyrLeu 49  
Db 445 TTAGTTGGATTTCGGATATTG 468  
RESULT 86  
A48779  
LOCUS Sequence 6 from Patent WO9604388.  
DEFINITION A48779.1 GI:2302448  
VERSION A48779.1  
KEYWORDS unclassified.  
SOURCE unclassified.  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 1149)  
AUTHORS Browne,M.J., Murphy,K.E., Chapman,C.G., Clinkenbeard,H.E., Young,P.R. and Shatzman,A.R.  
TITLE NOVEL COMPOUNDS  
JOURNAL Patent: WO 9604388-A 6 15-FEB-1996;  
SMITHKLINE BEECHAM PLC (GB)  
COMMENT Other publication AU 3382595 960304.  
FEATURES Location/Qualifiers  
1..1149  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 293 a 344 c 302 g 210 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 119 Length: 1149  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 4.04% Indels: 0  
DB: 6 Gaps: 0  
US-09-966-880A-8 (1-198) x A48779 (1-1149)  
QY 176 LeuArgArgIleLeuLeuProLeu 183  
Db 1002 CTCGACGAGATCCTTCCTCCTA 1025  
RESULT 88  
A48779  
LOCUS Sequence 6 from Patent WO9604388.  
DEFINITION A48779.1 GI:2302448  
VERSION A48779.1  
KEYWORDS unclassified.  
SOURCE unclassified.  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 1149)  
AUTHORS Browne,M.J., Murphy,K.E., Chapman,C.G., Clinkenbeard,H.E., Young,P.R. and Shatzman,A.R.  
TITLE NOVEL COMPOUNDS  
JOURNAL Patent: WO 9604388-A 9 15-FEB-1996;  
SMITHKLINE BEECHAM PLC (GB)  
COMMENT Other publication AU 3382595 960304.  
FEATURES Location/Qualifiers  
1..1149  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 293 a 344 c 304 g 208 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 119 Length: 1149  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 4.04% Indels: 0  
DB: 6 Gaps: 0  
US-09-966-880A-8 (1-198) x A48782 (1-1149)  
QY 176 LeuArgArgIleLeuLeuProLeu 183  
Db 1002 CTCGACGAGATCCTTCCTCCTA 1025  
RESULT 88  
A48782  
LOCUS Sequence 6 from Patent WO9604388.  
DEFINITION A48782.1 GI:2302450  
VERSION A48782.1  
KEYWORDS unclassified.  
SOURCE unclassified.  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 1149)  
AUTHORS Browne,M.J., Murphy,K.E., Chapman,C.G., Clinkenbeard,H.E., Young,P.R. and Shatzman,A.R.  
TITLE NOVEL COMPOUNDS  
JOURNAL Patent: WO 9604388-A 9 15-FEB-1996;  
SMITHKLINE BEECHAM PLC (GB)  
COMMENT Other publication AU 3382595 960304.  
FEATURES Location/Qualifiers  
1..1149  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 293 a 344 c 302 g 210 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 119 Length: 1149  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 4.04% Indels: 0  
DB: 6 Gaps: 0  
US-09-966-880A-8 (1-198) x A48782 (1-1149)  
QY 176 LeuArgArgIleLeuLeuProLeu 183  
Db 1002 CTCGACGAGATCCTTCCTCCTA 1025  
RESULT 88  
A48782  
LOCUS Sequence 6 from Patent WO9604388.  
DEFINITION A48782.1 GI:2302450  
VERSION A48782.1  
KEYWORDS unclassified.  
SOURCE unclassified.  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 1149)  
AUTHORS Browne,M.J., Murphy,K.E., Chapman,C.G., Clinkenbeard,H.E., Young,P.R. and Shatzman,A.R.  
TITLE NOVEL COMPOUNDS  
JOURNAL Patent: WO 9604388-A 9 15-FEB-1996;  
SMITHKLINE BEECHAM PLC (GB)  
COMMENT Other publication AU 3382595 960304.  
FEATURES Location/Qualifiers  
1..1149  
/organism="unidentified"  
/db\_xref="taxon:32644"

TITLE Direct Submission  
JOURNAL Submitted (03-APR-2002) Berkeley Drosophila Genome Project, Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA 94720, USA

COMMENT Sequence submitted by:  
Berkeley Drosophila Genome Project  
Lawrence Berkeley National Laboratory  
Berkeley, CA 94720

This clone was sequenced as part of a high-throughput process to sequence clones from Drosophila Gene Collection 1 (Rubin et al., Science 2000). The sequence has been subjected to integrity checks for sequence accuracy, presence of a polyA tail and contiguity within 100 kb in the genome. Thus we believe the sequence to reflect accurately this particular cDNA clone. However, there are artifacts associated with the generation of cDNA clones that may have not been detected in our initial analyses such as internal priming, priming from contaminating genomic DNA, retained introns due to reverse transcription of unspliced precursor RNAs, and reverse transcriptase errors that result in single base changes. For further information about this sequence, including its location and relationship to other sequences, please visit our web site (<http://fruitfly.berkeley.edu>) or send email to [cdna@fruitfly.berkeley.edu](mailto:cdna@fruitfly.berkeley.edu).

FEATURES  
Source Location/Qualifiers  
1..1101  
/organism="Drosophila melanogaster"  
/strain="y: cn bw sp"  
/db\_xref="taxon:7227"  
misc\_feature 1..1091  
/note="sim4 alignment with AE003477 (63B6-63D1)"  
CDS 750..1010  
/note="Longest ORF"  
/codon\_start=1  
/product="RH52423p"  
/protein\_id="AAM11287.1"  
/db\_xref="GI:20151856"  
/translation="MFNTKMYSVRIWLGSVLLGPLLLAAIAVQGOEPASPVFQNH  
RTKEWNLNDNTFSDCKRRSVGFYADMEYNCQIFHMCDEY"  
BASE COUNT 343 a 213 c 240 g 305 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 115 Length: 1101  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 4.04% Indels: 0  
DB: 3 Gaps: 0  
US-09-966-880A-8 (1-198) x AY094934 (1-1101)  
QY 42 PheSerLeuAspPheGlyTyrLeu 49  
Db 445 TTAGTTGGATTTCGGATATTG 468  
RESULT 86  
A48779  
LOCUS Sequence 6 from Patent WO9604388.  
DEFINITION A48779.1 GI:2302448  
VERSION A48779.1  
KEYWORDS unclassified.  
SOURCE unclassified.  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 1149)  
AUTHORS Browne,M.J., Murphy,K.E., Chapman,C.G., Clinkenbeard,H.E., Young,P.R. and Shatzman,A.R.  
TITLE NOVEL COMPOUNDS  
JOURNAL Patent: WO 9604388-A 6 15-FEB-1996;  
SMITHKLINE BEECHAM PLC (GB)  
COMMENT Other publication AU 3382595 960304.  
FEATURES Location/Qualifiers  
1..1149  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 293 a 344 c 302 g 210 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 119 Length: 1149  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 4.04% Indels: 0  
DB: 6 Gaps: 0  
US-09-966-880A-8 (1-198) x A48779 (1-1149)  
QY 176 LeuArgArgIleLeuLeuProLeu 183  
Db 1002 CTCGACGAGATCCTTCCTCCTA 1025  
RESULT 88  
A48779  
LOCUS Sequence 6 from Patent WO9604388.  
DEFINITION A48779.1 GI:2302448  
VERSION A48779.1  
KEYWORDS unclassified.  
SOURCE unclassified.  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 1149)  
AUTHORS Browne,M.J., Murphy,K.E., Chapman,C.G., Clinkenbeard,H.E., Young,P.R. and Shatzman,A.R.  
TITLE NOVEL COMPOUNDS  
JOURNAL Patent: WO 9604388-A 9 15-FEB-1996;  
SMITHKLINE BEECHAM PLC (GB)  
COMMENT Other publication AU 3382595 960304.  
FEATURES Location/Qualifiers  
1..1149  
/organism="unidentified"  
/db\_xref="taxon:32644"

```
antagonists and immunoglobulins
JOURNAL Patent: US 5783181-A 6 21-JUL-1998;
FEATURES
  source
    Location/Qualifiers
      1..1149
        /organism="unknown"
BASE COUNT 293 a 344 c 302 g 210 t
ORIGIN

Alignment Scores:
Pred. No.: 119 Length: 1149
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.04% Indels: 0
DB: 6 Gaps: 0

US-09-966-880A-8 (1-198) x AR018706 (1-1149)
Oy 176 LeuArgArgIleLeuLeuProLeu 183
Db 1002 CTCGACGGATCCTTCTCTCTA 1025

RESULT 89
LOCUS AR018708 1149 bp DNA linear PAT 05-DEC-1998
DEFINITION Sequence 9 from patent US 5783181.
ACCESSION AR018708
VERSION AR018708.1 GI:3973822
KEYWORDS
SOURCE
  ORGANISM
    Unclassified.
  REFERENCE
    1 (bases 1 to 1149)
    Brown,M.Joseph., Young,P.Ronald., Shatzman,A.Richard.,
    Murphy,K.Elizabeth., Chapman,C.Gerald. and
    Clinkenbeard,H.Elizabeth.
  TITLE
    Therapeutic uses of fusion proteins between mutant IL 4/IL13
    antagonists and immunoglobulins
  JOURNAL
    Patent: US 5783181-A 9 21-JUL-1998;
  FEATURES
    source
      Location/Qualifiers
        1..1149
          /organism="unknown"
BASE COUNT 293 a 344 c 304 g 208 t
ORIGIN

Alignment Scores:
Pred. No.: 119 Length: 1149
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.04% Indels: 0
DB: 6 Gaps: 0

US-09-966-880A-8 (1-198) x AR018708 (1-1149)
Oy 176 LeuArgArgIleLeuLeuProLeu 183
Db 1002 CTCGACGGATCCTTCTCTCTA 1025

RESULT 90
LOCUS A58738 1188 bp DNA linear PAT 06-MAR-1998
DEFINITION Sequence 1 from Patent WO9700319.
ACCESSION A58738
VERSION A58738.1 GI:3714280
KEYWORDS
SOURCE
  ORGANISM
    unidentified.
  REFERENCE
    1 (bases 1 to 1188)
    Brown,M.J., Chapman,C.G., Clinkenbeard, Helen,E. and Robinson,J.H.
    CHIMERIC LEPTIN FUSED TO IMMUNOGLOBULIN DOMAIN AND USE
    Patent: WO 9700319-A 1 03-JAN-1997;
  JOURNAL
    SMITHKLINE BEECHAM PLC (GB)
    Other publication AU 6011096 970115.
  FEATURES
    source
      Location/Qualifiers
        1..1188
          /organism="unidentified"
          /db_xref="taxon:32644"
BASE COUNT 282 a 369 c 311 g 226 t
ORIGIN

Alignment Scores:
Pred. No.: 123 Length: 1188
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.04% Indels: 0
DB: 6 Gaps: 0

US-09-966-880A-8 (1-198) x A58740 (1-1188)
Oy 176 LeuArgArgIleLeuLeuProLeu 183
Db 1044 CTCGACGGATCCTTCTCTCTA 1067

RESULT 91
LOCUS A58740 1188 bp DNA linear PAT 06-MAR-1998
DEFINITION Sequence 3 from Patent WO9700319.
ACCESSION A58740
VERSION A58740.1 GI:3714282
KEYWORDS
SOURCE
  ORGANISM
    unidentified.
  REFERENCE
    1 (bases 1 to 1188)
    Brown,M.J., Chapman,C.G., Clinkenbeard, Helen,E. and Robinson,J.H.
    CHIMERIC LEPTIN FUSED TO IMMUNOGLOBULIN DOMAIN AND USE
    Patent: WO 9700319-A 3 03-JAN-1997;
  JOURNAL
    SMITHKLINE BEECHAM PLC (GB)
    Other publication AU 6011096 970115.
  FEATURES
    source
      Location/Qualifiers
        1..1188
          /organism="unidentified"
          /db_xref="taxon:32644"
BASE COUNT 282 a 369 c 311 g 226 t
ORIGIN

Alignment Scores:
Pred. No.: 123 Length: 1188
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.04% Indels: 0
DB: 6 Gaps: 0

US-09-966-880A-8 (1-198) x A58740 (1-1188)
Oy 176 LeuArgArgIleLeuLeuProLeu 183
Db 1044 CTCGACGGATCCTTCTCTCTA 1067

RESULT 92
LOCUS AF238055 1230 bp DNA linear PLN 12-JUN-2001
DEFINITION Hydrastis canadensis NADH dehydrogenase subunit F (nuhf) gene,
partial cds; chloroplast gene for chloroplast product.
ACCESSION AF238055
VERSION AF238055.2 GI:10710339
KEYWORDS
SOURCE
  ORGANISM
    Hydrastis canadensis.
    Chloroplast Hydrastis canadensis
    Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
    Spermatophyta; Magnoliophyta; eudicotyledons; Ranunculales;
    Ranunculaceae; Hydrastis.
```



```

/db_xref="GI:3243077"
/translation="MIHTNSARSSFEASLSPSRRTLRLARLAPVEGPTVAELRAALKR
DEFTAPQPVVDACGLSGVATTIQNAHRHNGLYPEHCYMAALYRONLHIALAQQLL
AQVFDLAAMRRGNRHVGLFIQPEPICLSKRSDDVLRNLDHRGLPFAITLAVSPS
LLQSPYQQSLHSLRRLHAGCNLVLTSTFDESPHLDALEICGIALPELMYALR
QGGRAACRAIAQAQAQRIPVMAAGVDEMAQAADLLPCRYLLGDHVPMTGOR
FLYWYRREPPNP"

```

BASE COUNT 229 a 514 c 461 g 261 t  
ORIGIN

## Alignment Scores:

Pred. No.: 147 Length: 1465  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservativeness: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 4.04% Indels: 0  
DB: 1 Gaps: 0

US-09-966-880A-8 (1-198) x AF071567 (1-1465)

Qy 126 LeuAqrgrLeuHisArgAlaGly 133

Db 1066 CTGCGGCGCTGCACCGGCGGA 1089

## RESULT 95

EANDHF1

LOCUS

DEFINITION Exacum affine NADH dehydrogenase subunit (ndhf) gene, partial cds; Chloroplast gene for chloroplast product.

ACCESSION AF147710

VERSION AF147710.1 GI:6110314

KEYWORDS 1 of 2

SEGMENT Exacum affine.

SOURCE

ORGANISM Chloroplast Exacum affine

REFERENCE 1 (bases 1 to 1565)  
AUTHORS Olmstead,R.G., Kim,K.J., Jansen,R.K. and Wagstaff,S.J.  
TITLE The phylogeny of the Asteridae sensu lato based on chloroplast ndhF gene sequences

JOURNAL Mol. Phylogenet. Evol. 16 (1), 96-112 (2000)

MEDLINE 20340987  
PUBMED 10877943

REFERENCE 2 (bases 1 to 1565)

AUTHORS Olmstead,R.G., Kim,K.-J., Jansen,R.K. and Wagstaff,S.J.

TITLE Direct Submission

JOURNAL Submitted (03-MAY-1999) Botany, University of Washington, Campus Box 355325, Seattle, WA 98195, USA

FEATURES Location/Qualifiers

source 1..1565

/organism="Exacum affine"

/organelle="plastid:chloroplast"

/db\_xref="taxon:13525"

<1..>1565

/gene="ndhf"

/codon\_start=2

/product="NADH dehydrogenase subunit"

/protein\_id="AA03769.1"

/db\_xref="GI:6110322"

/translation="WIIFPLPVPFAGLLLPATKSFRRMAFQSVLLISIVM

XFSYILSIQIINSSFFQYVWSMTINNDFSLDFGLDPLTSTMSITITVGLMWLIY

SDNYMDQCYLRFAYVMSFSTSMGLAVTSSNLIQIYIFWELVGLCSYLLIGFWFTR

PVANAACQAFVNRVDFGLLGLGYFITGSFSDRLFEINLLIYNNVNFVLFV

TLCAVLDFAGAVAKSAQFPDHWLPPDAMEGTPISALHAATWAAGIFLVARLLPLF

RVIPYINLYSGIITVIGLGTALALAKDKIKRGLAYSTMSQLGMMALGMSYRNA

## Alignment Scores:

Pred. No.: 156 Length: 1565  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservativeness: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 4.04% Indels: 0  
DB: 8 Gaps: 0

US-09-966-880A-8 (1-198) x EANDHF1 (1-1565)

Qy 42 PheSerLeuAspPheGlyTyrLeu 49

Db 218 TTTTCTTTAGACTTCGGATACTTG 241

## RESULT 96

AF353205

LOCUS

DEFINITION Azospirillum brasilense strain Sp7 poly-beta-hydroxybutyrate synthase (phbc) gene, complete cds.

ACCESSION AF353205

VERSION AF353205.1 GI:21326930

KEYWORDS

SOURCE

ORGANISM Azospirillum brasilense.

Bacteria; Proteobacteria; alpha subdivision; Rhodospirillaceae;

Azospirillum.

REFERENCE 1 (bases 1 to 1857)

AUTHORS Kadouri,D., Burdman,S., Jurkevitch,E. and Okon,Y.

TITLE Identification and Isolation of Genes Involved in

Poly(beta-Hydroxybutyrate) Biosynthesis in Azospirillum brasilense

and Characterization of a phbc Mutant

JOURNAL Appl. Environ. Microbiol. 68 (6), 2943-2949 (2002)

MEDLINE 22034968

PUBMED 12039753

REFERENCE 2 (bases 1 to 1857)

AUTHORS Kadouri,D.E., Jurkevitch,E. and Okon,Y.

TITLE Direct Submission

JOURNAL Submitted (25-FEB-2001) Department of Plant Pathology and

Microbiology, The Hebrew University of Jerusalem, Faculty of

Agricultural, Food and Environmental Quality Sciences, Rehovot

76100, Israel

## FEATURES

source

1..1857

/organism="Azospirillum brasilense"

/strain="Sp7"

/db\_xref="taxon:192"

1..1857

/gene="phbc"

1..1857

/gene="phbc"

/codon\_start=1

/transl\_table=11

/product="poly-beta-hydroxybutyrate synthase"

/protein\_id="AA048100.1"

/db\_xref="GI:21326931"

/translation="MYGDLGVDDAQFTORLLDLVLKLPFGVGAAGDALHFAHLLLLD

LVRVGEERFLSRQADGVGAKNPDPMGVGHAFLEMTTRMADPAKIMKAOMTLWDY

LTLMQRTQRFQGDAPQIPAPAKDDRRFKDSADMENTLDFDIKOSYLLSARWMTV

NEVDGLDDHTAKRVDYTRQFVDAMAPSNFVMTNPEVLTRTTIEGGENLVKGLHLLK

DLBERGKELRISMTDYDAFOGVKNIAVTGKVVFTDLMLQ1QVTPPTPEVKNRPLMI

LPWINKYIYILDREKNSFIKWAVDGCHSVFLSWVNPDEK1AQKGFEDYMFECVLA

LDALKEVGTGKDYNALCYCLGCTLLASTLSYMAAKKDDR1KSATFFTMJDFTFACFL

SVTIDEQLTWIESQAAQOGLGSKMATTFNMLRANDLIWSFVNNYILCKDFFPFD

LLYNSDSTRMPAAMHSFYLRNMYOKNLLAOPGAVTLGGVPI1DIRKVKTPSPFLSAHE

DHAPKSTYMGAAHFLSGPVKFLVLAASGHIAGVVNPVPAACKYCYWTNAKLPKASDML

ASSEQTPGSMWPENNWVSTFSECKVPARNPEKGGPLVEDAPGSYAKVRIV"

BASE COUNT 357 a 613 c 575 g 312 t

ORIGIN

## Alignment Scores:

Pred. No.: 181 Length: 1857  
Score: 8.00 Matches: 8

```

Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.04% Indels: 0
DB: 1 Gaps: 0

US-09-966-880A-8 (1-198) x AF353205 (1-1857)

QY 123 ProGluGlyLeuArgArgLeuHis 130 linear PRI 15-JUL-2002
LOCUS AK094292 1886 bp mRNA BRACE2006249.
DEFINITION Homo sapiens cDNA FLJ36973 fis, clone BRACE2006249.
ACCESSION AK094292
VERSION AK094292.1 GI:21753322
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens cerebellum cDNA to mRNA, clone_lib:BRACE2
clone:BRACE2006249.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
AUTHORS Tanigami,A., Fujiwara,T., Shibahara,T., Goto,Y., Hirao,M.,
Shimizu,F., Wakebe,H., Ono,T., Hishigaki,H., Watanabe,T., Ozaki,K.,
Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S.,
Yanamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T.,
Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M.,
Kikuchi,H., Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K.,
Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y.,
Sugano,S., Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
NEDO human cDNA sequencing project
Unpublished
JOURNAL 2 (bases 1 to 1886)
REFERENCE Isogai,T. and Yanamoto,J.
AUTHORS Direct Submission
JOURNAL Submitted (04-JUL-2002) Takao Isogai, FLJ Project(HRI Team): 2-6-7-
Kazusa-Kanatairi, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: HRI and RAB.
FEATURES
source
Location/Qualifiers
1..1886
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="BRACE2006249"
/tissue_type="cerebellum"
/clone_lib="BRACE2"
/note="cloning vector: pME18SFL3"
BASE COUNT 540 a 360 c 335 g 651 t
ORIGIN
Alignment Scores:
Pred. No.: 183 Length: 1886
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.04% Indels: 0
DB: 9 Gaps: 0

US-09-966-880A-8 (1-198) x AK094292 (1-1886)

QY 97 PheLeuArgGlyAsnProAsnLeu 104
LOCUS TTTCTTGGGGCAATCCAACTTA 1039
DEFINITION 1016 TTTCTTGGGGCAATCCAACTTA 1039
Goyazia rupicola NADH dehydrogenase (ndhF) gene, partial cds;
chloroplast gene for chloroplast product.

RESULT 98
HAU12660 1938 bp DNA linear PLN 19-APR-2002
LOCUS HAU12660
DEFINITION Hemigraphis alternata NADH dehydrogenase subunit (ndhF) gene,
partial cds; chloroplast gene for chloroplast product.
ACCESSION U12660
VERSION U12660.1 GI:607930
KEYWORDS
SOURCE Hemigraphis alternata.
ORGANISM Chloroplast Hemigraphis alternata
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Lamiales; Acanthaceae; Hemigraphis.
1 (bases 1 to 1938)
Higher-level systematics of Acanthaceae determined by chloroplast
DNA sequences
Am. J. Bot. 82 (2), 266-275 (1995)
REFERENCE 2 (bases 1 to 1938)
AUTHORS Scottland,R.W., Sweere,J.A., Reeves,P.A. and Olmstead,R.G.
TITLE Direct Submission
JOURNAL Olmstead,R.G.
AUTHORS Submitted (26-JUL-1994) Richard G. Olmstead, E.P.O. Biology,
JOURNAL University of Colorado, Boulder, CO 80309, USA
FEATURES
source
Location/Qualifiers
1..1938
/organism="Hemigraphis alternata"
/organelle="plastid:chloroplast"
/db_xref="taxon:34269"
1..1938
/gene="ndhF"
/gene="ndhF"
/note="151 nucleotides from the PCR primer near the 5' end
were unsequenced"
/codon_start=1
/trans_table=11
/product="NADH dehydrogenase subunit"
/protein_id="AAAG1724.1"
/db_xref="GI:607931"
VGI:AVLYSDNYMAHDGGLRFFAYMSFFSTMLGLVTSNLIQIYIPWELVGCSTYL
LIGFVTPRLAADCQAKAFV7NRVDFGLLGLIFCLFWITSGFRDLUFEMFNNSVYN
NQVNFPTLCADLVAFAGAKSAQFPLHWLPDAMEGPTISALIHAATVWAGVFL
VARLLPFWIPYIMDFISLIGIITFLGATLALAKDKIKRGLAYSTMQLGYYMLAL
TKGYSRGLAFHLTHAYSKALLFLGSGSVIHSMESIVGSPDKQNIIVLMGGLRHLPL
ITKTSILGLTFLCGVPLACFWKDEILNDTWLYSPIFATIAITAGLTXYMFRIY
LLTEGHLNVHFQNSANQIFLYSISLWACSKKININFRLLRINKNTSFLLKTP
ORDEKINRPFINILSFHNKTSFYTESDNTMLFLLILVLTFSFVGSIGIPNPF
NOEGTDILSKWVTPSIHLLHQAEKSTHWYEFKDATFSVSLAYFGIFLAALLKPIY
SSKFNFDLVNSFVKTNLRDWDKRVINALYNWYNRAYIDVFYRTSVIAMIRGLAKLT
HFVDRRIIDGITNLVGLLSXFI"
BASE COUNT 535 a 303 c 342 g 756 t 2 others
ORIGIN
Alignment Scores:
Pred. No.: 187 Length: 1938
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.04% Indels: 0
DB: 8 Gaps: 0

US-09-966-880A-8 (1-198) x HAU12660 (1-1938)

QY 42 PheSerLeuAspPheGlyTyrLeu 49
LOCUS TTTCTTTAGACTTCGGCTACTTG 90
DEFINITION 67 TTTCTTTAGACTTCGGCTACTTG 90
PheSerLeuAspPheGlyTyrLeu 49
RESULTS 99
AF257485 1990 bp DNA linear PLN 11-JUN-2000
LOCUS AF257485
DEFINITION Goyazia rupicola NADH dehydrogenase (ndhF) gene, partial cds;
chloroplast gene for chloroplast product.

```

```
ACCESSION AF257485
VERSION AF257485.1 GI:8453157
KEYWORDS
SOURCE
ORGANISM
Chloroplast Goyazia rupicola
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots;
Asteridae; euasterids I; Lamiales; Gesneriaceae; Goyazia.
REFERENCE 1 (bases 1 to 1990)
AUTHORS Smith, J.F.
TITLE The phylogenetic relationships of Lemnocarpaceae and Goyazia
(Gesneriaceae) based on ndhF sequences
JOURNAL Ann. Mo. Bot. Gard. (2000) In press
REFERENCE 2 (bases 1 to 1990)
AUTHORS Smith, J.F.
TITLE Direct Submission
JOURNAL Submitted (18-APR-2000) Biology, Boise State University, 1910
University Drive, Boise, ID 83725, USA
FEATURES
source
1..1990
/organism="Goyazia rupicola"
/organella="plastid:chloroplast"
/db_xref="taxon:125973"
<1..>1990
/gene="ndhF"
<1..>1990
/gene="ndhF"
/codon_start=1
/product="NADH dehydrogenase"
/protein_id="AAF75262.1"
/db_xref="GI:8453158"
/translation="GMIIPILPVPMLIGTGLLLPPTATKRLRMALPSILLSTV
MFSINLSIOLINSIIQVWSTXDNDSDGFLYDILLTSTMSLITTVGIMVLI
YSDNYMAHDGGLRFLVYMFSEFSGFLVTSNLIQIYLFWELVGCVSILLGFNET
RPLAANCQAFVYNNVGDGGLLIGISFTWIGSFEFRDLFEILLNNEVDFFP
VTLCAALLFAGVAKSVQFPLHVLDPAMEGPTPSALIIHAATWAAAGIFLVARLPL
FIATPYIMYSIVVGIITVLGATLALAKQDKRGLAYSAMSOGLYMMALMGSYRC
ALPHLIHAYSKALLFLGSGSIHSMETIVGSPDKSQNMGLMGLTKHVPIYKISFL
FGTSLGCIPLACFKWDEILNDSWLSPIFAIASATAGTAVTAFTYRIYLLTFEGH
LNIHFONYSGKONSPIYSISIWGKXXKXKRLNDNNEXXTFFFAKTKYRSDDKGIK
MARFFINIVHDFKKPSYIESDNTMLFLLVLLVFTLVGSLGIFPNXXEXLDLIL
TKWLAPINILLHOKNSDMWIEFFKGAVSVSIAIXGIFVASFLYKPIYSSSKNFEL
INSFVKGQPSILMDKTINGYDMSYNRAVIDTFYTRSF"
BASE COUNT 542 a 313 c 346 g 770 t 19 others
ORIGIN
Alignment Scores:
Pred. No.: 192 Length: 1990
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.04% Indels: 0
DB: 8 Gaps: 0
US-09-966-880A-8 (1-198) x AF257485 (1-1990)
Qy 42 PheSerLeuAspPheGlyTyrLeu 49
|||||
Db 220 TTTTCTTTAGATTTCGGCTACTTC 243
RESULT 100
LOCUS
DEFINITION Eremomastax speciosa NADH dehydrogenase subunit (ndhF) gene,
partial cds; Chloroplast gene for chloroplast product.
ACCESSION U12659
VERSION U12659.1 GI:607928
KEYWORDS
SOURCE
ORGANISM
Eremomastax speciosa
Chloroplast Eremomastax speciosa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Lamiales; Acanthaceae; Eremomastax.
REFERENCE 1 (bases 1 to 2025)
AUTHORS Scotland, R.W., Sweere, J.A., Reeves, P.A. and Olmstead, R.G.
TITLE Higher-level systematics of Acanthaceae determined by chloroplast
DNA sequences
JOURNAL Am. J. Bot. 82 (2), 266-275 (1995)
REFERENCE 2 (bases 1 to 2025)
AUTHORS Olmstead, R.G.
TITLE Direct Submission
JOURNAL Submitted (26-JUL-1994) Richard G. Olmstead, E.P.O. Biology,
University of Colorado, Boulder, CO 80309, USA
FEATURES
source
1..2025
/organism="Eremomastax speciosa"
/organella="plastid:chloroplast"
/db_xref="taxon:37794"
1..2025
/gene="ndhF"
<1..>2025
/gene="ndhF"
/note="70 nucleotides from the PCR primer near the 5' end
were unsequenced"
/codon_start=1
/transl_table=11
/product="NADH dehydrogenase subunit"
/protein_id="AA61723.1"
/db_xref="GI:607929"
/translation="ATKNRLRLWTPTXXLSIVMIFSTKLSIEQINGSSIIYOYVXW
TFDNDPSLDFGLYLDPLTSLMLITTVGTLITVLIYSDNYNAHDGFLRFPAYMSFCT
AMEGLVTSNLIIQIYIFWELVGCVSILLIGFTVRLPLAADACQAFVNNRVGDFGLLL
GIFCFWITGTFEFDLRFENFNSVNNVNFVTLCTVLLFAGAVAKSAQPLHWW
LPDAMEGPTPSALIIHAATWAAAGVFLVARLLPLFWIPYIMDIISIIGLITLLGAT
LALAQDIDIKRLAYSTMSOLGYMMALGMCSCRSALFHLITHAYSKALLFLGSGSVIH
SMESIVGSPDKNQNMGLMGLTKHLPTIKIAELLGTFLGCGVPLACFPWSKDEILNH
TWLSPSFAIIASITAGLTAFYMFRIYLLTTFEGHNVHLNLYSANOINIFLYSISLMCK
ACSKRINFNFSLRIYNTSSSLKKTQORDEKRRGMFPFKILYSRNNKQTFSTYPE
SDNTMLFSLILVFTLVGSGIGIPFNENQEGTDLISKWVTDSIILHQKAKDSTHYE
YEFKDATFSVSLAYTGIFLASSLYKPIYSSFKNFDLINSFVKTGLNRRDWMKMINVL
YNNWSNRGYIDTFYITISFFAMIRGLAKLTHFVDRRIIDGITNVVGLMSPFI"
BASE COUNT 557 a 310 c 360 g 788 t 10 others
ORIGIN
Alignment Scores:
Pred. No.: 195 Length: 2025
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 8 Gaps: 0
US-09-966-880A-8 (1-198) x ESU12659 (1-2025)
Qy 42 PheSerLeuAspPheGlyTyrLeu 49
|||||
Db 148 TTTTCTTTAGACTTCGGTTACTTC 171
Search completed: July 17, 2003, 22:23:21
Job time : 2445 secs
```